

BAYESIAN ANALYSIS OF COMPETING RISKS MODELS

By

CHEN-PIN WANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1999

© Copyright 1999

by

Chen-Pin Wang

To my family

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Dr. Malay Ghosh, without whom this work would never have been completed. He was always willing to give me guidance, knowledge, and friendship when it was needed. I would also like to thank my committee members, Dr. Randolph Carter, Dr. Andrew Rosalsky, Dr. Mark Yang, and Dr. Yumei Chen, for supporting my research. Additionally, the collective knowledge and experience of my fellow students and members of the faculty never failed to inspire and encourage me during the completion of my work. I would like to especially thank Dr. Roman Littel for being willing to help me beyond statistics.

Finally, I would like to thank my family, for their love and support, both financial and emotional. I would especially like to thank my best friend in Gainesville, F. J. Huang, who put up with me and constantly encouraged me during the completion of my doctoral work, without whom this work would have never been finished.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	ix
CHAPTERS	
1 LITERATURE REVIEW	1
2.1 Introduction	1
2.2 Competing Risks	2
2.3 Bivariate Lifetime Distributions	5
2.4 Noninformative Priors	19
2.5 Research Proposal	23
2 BAYESIAN ANALYSIS OF SELECTED BIVARIATE EXPONENTIAL MODELS	25
2.1 Introduction	25
2.2 Notation	26
2.3 Noninformative Prior Modification	27
2.4 Bayesian Analysis for Marshall-Olkin BVE	29
2.5 Bayesian Analysis for the ACBVE Models	32
2.6 Prior Performance	37
2.7 Identifiable ACBVE via Informative Priors	48
3 BAYESIAN ANALYSIS OF GENERALIZED BIVARIATE EXPONENTIAL MODELS	51
3.1 Introduction	51
3.2 Generalized BVE Models	52
3.3 Generalization of ACBVE Models	53
3.4 Generalization of the Marshall-Olkin BVE Model	59
3.5 Application to a Sample with Categorical Covariates	62

3.6	Sampling Schemes	65
3.7	Illustrated Examples	67
4	GEOMETRIC COMPETING RISKS MODELS	78
4.1	Introduction	78
4.2	Generalized Geometric Models and Likelihood Functions	79
4.3	Bayesian Analysis	83
4.4	Application to a Sample with Categorical Covariates	88
4.5	Data Analysis	90
5	SUMMARY AND FUTURE RESEARCH	94
APPENDIX		
	PROOFS OF MATCHING PROPERTIES OF π_{JU} AND π_J	95
	REFERENCES	96
	BIOGRAPHICAL SKETCH	100

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1 Reference Priors for the Marshall-Olkin BVE	31
2.2 Posterior Distributions for the Marshall-Olkin BVE	32
2.3 The ACBVE Reparameterization	33
2.4 Posterior Distributions for λ and ρ of ACBVE's	36
2.5 Posterior Distributions (Moments) for ϕ of ACBVE's	36
2.6 Frequentist Coverage Probabilities for λ under $\pi_U, \pi_{UU}, \pi_J, \pi_{JU}$	39
2.7 Numerical Simulation Results of Table 2.6	39
2.8 Posterior Estimates for ϕ of ACBVE's under Beta Priors	49
3.1 Probability Matching for $\mathbf{A}^{-1}(1, 0.5)'$ under $\pi_{JU}, \pi_{UU}, \pi_J, \pi_U$	68
3.2 Probability Matching for $\mathbf{A}^{-1}(1, 0.5, 2)'$ under $\pi_{JU}, \pi_{UU}, \pi_J, \pi_U$	69
3.3 ECOG Posterior Estimates under Main Effect Model and π_{UU}	73
3.4 ECOG Posterior Estimates under Interaction Model and π_{JU}, π_{UU}	75
3.5 Posterior Estimates for Contrasts of Interest under π_{UU} (or π_{JU})	77
4.1 Stochastic Transition Status	80
4.2 Comparison of Posterior Estimates for $P(\Delta = 1 T^{(L)} = 1)$	90

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 Transformed Q-Q Plots for ϕ under π_J , π_U , and $n = 10$	42
2.2 Transformed Q-Q Plots for ϕ under π_J , π_U , and $n = 50$	43
2.3 Transformed Q-Q Plots for ϕ under π_J , π_U , and $n = 100$	44
2.4 Transformed Q-Q Plots for ρ under π_J , π_U , and $n = 10$	45
2.5 Transformed Q-Q Plots for ρ under π_J , π_U , and $n = 50$	46
2.6 Transformed Q-Q Plots for ρ under π_J , π_U , and $n = 100$	47
2.7 Beta Priors and the Associated Posteriors of ϕ under $n = 10, 50$	50
3.1 R's and Posterior Densities of β_1 , β_2 , β_3 , and β_4 under π_{UU}	71
3.2 R's and Posterior Densities of γ_1 , γ_2 , γ_3 , and γ_4 under π_{UU}	72
3.3 Analysis 1: Goodness-of-fit Plots for T_i 's and Δ_i 's under BVE	74
3.4 Analysis 2: Goodness-of-fit Plots for T_i 's and Δ_i 's under BVE	76
4.1 Bayesian Goodness-of-Fit Plot for $n_{1j}^{(l)}$'s	91
4.2 Survival Curves for T under the ACBVE and the GEM	92

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

BAYESIAN ANALYSIS OF COMPETING RISKS MODELS

By

Chen-Pin Wang

August 1999

Chairman: Malay Ghosh
Major Department: Statistics

Bivariate exponential (BVE) models have been widely used in the analysis of competing risks data involving two risk components. For such analysis, frequentist approach often runs into difficulty due to nonidentifiable likelihood. With an end to overcome the nonidentifiability, recent literature has been geared towards Bayesian analysis with informative priors. However, systematic prior elicitation is often difficult. This study focuses instead on Bayesian analysis with noninformative priors.

Due to the nature of the competing risks data and BVE model structure, quite often it leads a likelihood function with a nonregular Fisher information matrix impeding thereby the calculation of standard noninformative priors such as Jeffreys' priors and their variants. As a remedy, a stagewise noninformative prior elicitation strategy is proposed. A variety of noninformative priors are developed, and are used for data analysis. In addition, the frequentist probability matching criteria are investigated among the newly developed noninformative priors.

Often due to resource limitation or other economic and practical reasons, individuals are only periodically screened. The resulting competing risks data are necessarily discrete. A class of flexible models is introduced to handle such discretized data.

CHAPTER 1 LITERATURE REVIEW

1.1 Introduction

Often in a life-testing situation, failure of an individual can be identified as one or more of s ($s > 1$) mutually exclusive, but possibly dependent causes of failure. In other words, each individual is subject to s distinct risks referred to as *competing risks* threatening its life. Associated with cause i , there is a nonnegative absolutely continuous random variable T_i representing the lifetime of an individual when no other potential risks are present. Suppose that the termination time of an individual is defined as the time to the first failure. Thus, lifetime of an individual is given by $T = \min\{T_1, \dots, T_s\}$. The available information is usually given by the pair (T, I) , where I indicates the cause(s) of failure. The competing risks concept can appropriately be applied to many areas of study, such as industrial reliability analysis, market transaction analysis, and clinical trial on paired organs.

Our main goal is to use Bayesian methodology for making statistical inference in competing risks models to study certain lifetime features of interest and the covariate effects on the underlying survival functions. The first step is necessarily multivariate lifetime modeling. In the one-dimensional case, the Weibull distribution has been considered as the most flexible one for lifetime modeling since it accommodates three major types of failure rates: aging type, decaying type, and constant type. A natural extension to multidimensional lifetime modeling would be the multivariate Weibull model. An important special case is the multivariate exponential distribution. When the shape parameters are known, then one can make a power transformation on

Weibull lifetimes to bring the resulting distributions into the exponential form. Our primary emphasis will be on the two-dimensional case. As is well-known, typical bivariate exponential distributions are generated from Poisson processes. Some are designated for distinctive characteristics of lifetimes, such as memorylessness and absolute continuity (Freund (1961) and Sarkar (1987)). There exists an inevitable tradeoff between these properties except when the two variables are independent (Block and Basu, 1974). Also, practical considerations have often led to bivariate exponential distributions. For example, Ryu's (1993) shock model is derived purely from practical considerations.

The organization of this chapter is as follows. In Section 1.2, we introduce the concept and applications of competing risks. Also, one will be reviewing some relevant work in this area. A detailed discussion of bivariate lifetime models will be illustrated through examples in Section 1.3. Formulation of noninformative priors will be reviewed in Section 1.4. In Section 1.5, our objectives for studying competing risks models with Bayesian methodology are proposed.

1.2 Competing Risks

1.2.1 Concepts and Applications

Consider a study of an in-processing population. Independent individuals of the population are exposed to s mortality sources that cause termination of the current process (e.g., life, presence of certain feature), namely, failure. Defined in general, the term “lifetime” is the time length between entry of the study and the time of termination. Lifetime of each individual hence is subject to the s risk factors which are referred to as “competing risks”. This concept is applied, for example, in a series system consisting of s components, where failure of any one of the components leads to failure of the whole system. For convenience, we shall often refer to the

s -component series system as an illustrative example. As an example of animate application, a clinical trial is considered, where the effect time of a new treatment for a certain disease on pairs of organs measures the time difference between the cure and the time to the first infection of the diagnosed organ(s). In the simplest competing risks setting, *coherence*, independence, and mutual exclusiveness are assumed among the possible risks. *Coherence* is our major assumption, meaning that individuals experience simultaneous exposures to all possible risk factors throughout the study. Coherence is only convenient for counting lifetime. The mutual exclusiveness can be achieved by refining the classification of the risk factors. However, the independence assumption may not always be appropriate. If this is the case, then one must consider multivariate lifetime distributions. Our primary interest in competing risks context are: the survival or failure rate in the presence of all risk factors and the cause-specific probabilities. To this end, we need the knowledge of the distribution of (T, I) .

1.2.2 Models and Likelihood Functions

Let $\boldsymbol{\theta}$, $L(\cdot)$, $P(\cdot)$, and $p(\cdot)$, and $S(\cdot)$ denote model parameter vector, likelihood function, probability measure, probability density, and survival probability measure respectively. The joint probability density function of $(T_1, \dots, T_s)$ is denoted by $f(t_1, \dots, t_s | \boldsymbol{\theta})$, and the joint survival function $\int_{t_1}^{\infty} \dots \int_{t_s}^{\infty} f(y_1, \dots, y_s | \boldsymbol{\theta}) dy_1, \dots, dy_s$ is denoted by $S(t_1, \dots, t_s | \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Let $T = \min\{T_1, \dots, T_s\}$, while I identifies the cause of failure. Suppose that $(T, I) = (t, i)$ is observed. Then the joint density $p(t, i | \boldsymbol{\theta})$ is given by

$$p(t, i | \boldsymbol{\theta}) = \int_t^{\infty} \dots \int_t^{\infty} f(t_1, t_2, \dots, t_{i-1}, t, t_{i+1}, \dots, t_s | \boldsymbol{\theta}) \prod_{j \neq i} dt_j. \quad (1.1)$$

Alternatively, (1.1) can be rewritten as

$$p(t, i | \boldsymbol{\theta}) = \left[-\frac{\partial S(t_1, \dots, t_s | \boldsymbol{\theta})}{\partial t_i} \right] \Big|_{t_1=t_2=\dots=t_s=t}.$$

Under the independence of $T_1, \dots, T_s$, (1.1) simplifies to

$$p(t, i|\boldsymbol{\theta}) = f_i(t|\boldsymbol{\theta}_i) \prod_{j \neq i} S_j(t|\boldsymbol{\theta}_j),$$

where f_i and S_i denote the marginal pdf and survival function for the i^{th} component, respectively. However, this assumption of independence usually is not practical.

For modeling dependence, Chiang (1961) suggested the constant proportional hazard rate model. Let $f(t|\boldsymbol{\theta})$ and $S(t|\boldsymbol{\theta})$ denote the marginal pdf and survival function of T , respectively. The model assumes that the proportions of component-specific instantaneous failure rates, $\{r_i(\boldsymbol{\theta}) \equiv r_i(t|\boldsymbol{\theta}) = p(t, i|\boldsymbol{\theta})/S(t|\boldsymbol{\theta}) : 1 \leq i \leq s\}$, are constant through time. As a mathematical result, the likelihood function is fully determined by $f(t)$ (or $S(t)$) and the proportion constants $\{r_1(\boldsymbol{\theta}), \dots, r_s(\boldsymbol{\theta})\}$, i.e.,

$$p(t, i|\boldsymbol{\theta}) = [f(t|\boldsymbol{\theta})]^{r_i(\boldsymbol{\theta})} \prod_{i \neq j} [S(t|\boldsymbol{\theta})]^{r_j(\boldsymbol{\theta})}. \quad (1.2)$$

The Weibull (including the exponential distribution as a special case) and the Gompertz distribution (equivalent to logarithm of a Weibull-truncated-at-1) are some of the distributions that have been widely considered for one-dimensional lifetime modeling.

Lagakos and Williams (1977) and Kalbfleisch and Mackay (1979) proposed the constant-sum model in the analysis of informative censored survival data. Under the rationale that censoring constitutes an additional source of risk by preventing the identification of any future failure, this development is also appropriate for competing risks modeling, where $s = 2$. Let T_1 and T_2 denote failure time and censored time, respectively. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, and $N_t = (t, t + \tau)$. Two quantities, $a(t|\boldsymbol{\theta}) = P(T_2 \in N_t, I = 2|T_2 \in N_t, \boldsymbol{\theta})$ and $du(t|\boldsymbol{\theta}) = P(T_1 \in N_t, I = 1|T_2 \geq t, \boldsymbol{\theta})$, are defined such that $a(t|\boldsymbol{\theta}) + u(t|\boldsymbol{\theta}) = 1$. With this, the joint pdf $p(t, i|\boldsymbol{\theta})$ is given by

$$\begin{aligned} p(t, i|\boldsymbol{\theta}) &= [a(t|\boldsymbol{\theta}) dF_2(t|\boldsymbol{\theta}_2)]^{1_{\{i=2\}}} [(1 - F_2(t|\boldsymbol{\theta}_2)) du(t|\boldsymbol{\theta})]^{1_{\{i=1\}}} \\ &= [a(t|\boldsymbol{\theta}) dF_2(t|\boldsymbol{\theta}_2)]^{1_{\{i=2\}}} [(1 - F_2(t|\boldsymbol{\theta}_2)) d(1 - a(t|\boldsymbol{\theta}))]^{1_{\{i=1\}}}, \end{aligned} \quad (1.3)$$

where d denotes the differentiation operator and $\mathbf{1}$ is the conventional indicator function.

Other alternatives, such as the conditional independence technique, have also shown their strength in the competing risks modeling. Motivated by Vaupel (1979), Clayton and Cuzick (1985) made use of the positive random quantity, *frailty*, to construct models for correlated lifetimes. Given the unobserved frailty, the true hazard functions act multiplicatively to the baseline hazards and they are conditionally independent. Suppose that the system frailty w is characterized by the pdf $\pi(w|\boldsymbol{\eta})$. Let $\{h_i(t|\boldsymbol{\theta}) = -\partial \log[S_i(t|\boldsymbol{\theta})]/\partial t : 1 \leq i \leq s\}$ be the marginal hazard rates and $\{H_i(t|\boldsymbol{\theta}) = -\log[S_i(t|\boldsymbol{\theta})] : 1 \leq i \leq s\}$ be the corresponding cumulative hazard rates. Then

$$p(t, i|\boldsymbol{\theta}, \boldsymbol{\eta}) = \int h_i(t|\boldsymbol{\theta}) w e^{-\sum_{i=1}^s H_i(t|\boldsymbol{\theta}) w} \pi(w|\boldsymbol{\eta}) dw. \quad (1.4)$$

In order to express the integral explicitly, one needs to specify the pdf $\pi(w|\boldsymbol{\eta})$ in addition to the underlying lifetime distribution. Gamma or positive stable distribution are some of the most commonly used choices.

From now on, we shall be considering exclusively the case $s = 2$. In the following section, the bivariate Weibull family will be reviewed. Naturally, its scope extends well beyond the analysis of competing risks data.

1.3 Bivariate Lifetime Distributions

The parametric modeling for a one-dimensional survival function $S(t|\boldsymbol{\theta})$ is usually determined by its instantaneous hazard rate $h(t|\boldsymbol{\theta})$, or by the cumulative hazard function $H(t|\boldsymbol{\theta})$ since $S(t|\boldsymbol{\theta}) = e^{-H(t|\boldsymbol{\theta})} \equiv e^{-\int_0^t h(u|\boldsymbol{\theta}) du}$. One familiar example is the constant hazard rate which gives rise to an exponential lifetime model. Its generalization to the two dimensional cases may be based on several considerations: absolute

continuity of the joint distribution, exponentiality of the marginals, exponentiality of the minimum, or positive (or negative) correlation between T_1 and T_2 .

Due to its flexibility, the bivariate Weibull (BVW) family has been widely applied for bivariate lifetime modeling. We especially incorporate it with the competing risks context. A bivariate distribution function is often referred to as a BVW if it meets at least one of the following conditions: (1.a) it has Weibull distributions marginally, e.g., Marshall-Olkin (1967) BVW, and (1.b) the minimal lifetime has Weibull marginal, e.g., Lu and Bhattacharyya's (1990) BVW.

We may observe that bivariate exponential distribution (BVE) is a special case of a BVW. Essentially, the BVE family constitutes the foundation of BVW. In the next subsection, we focus attention exclusively to the BVE subfamily.

1.3.1 Bivariate Exponential Distributions

A BVE can be generalized in several ways. Below, we discuss certain properties which are useful for such generalization.

1.3.1.1 Loss-of-Memory Property and Absolute Continuity

The very unique characterization for the exponential distribution is the *loss-of-memory property* (LMP). The interpretation for LMP is that given the current survival status, the joint distribution of the residual lifetimes remains the same as the unconditional joint lifetime distribution. For the univariate case, under assumed continuity of the distribution, this leads to an exponential distribution. In the multivariate case, this property is sometimes cited in a more specific way. Suppose that the property holds marginally (i.e. T_1 , T_2 , or T is marginally exponentially distributed). We refer to it as the *marginal loss-of-memory property* (MLMP). The term used to

describe LMP for a two-dimensional random vector, say (T_1, T_2) , is the *joint loss-of-memory property* (JLMP). It is defined by

$$S(t_1 + y, t_2 + y|\boldsymbol{\theta}) = S(t_1, t_2|\boldsymbol{\theta}) S(y, y|\boldsymbol{\theta}), \quad (1.5)$$

$\forall t_1 > 0, t_2 > 0$, and $y > 0$.

The family of bivariate distributions satisfying (1.5) has been characterized by Marshall and Olkin (1967). Let $S_1(t|\cdot) = P(T_1 > t|\cdot)$ and $S_2(t|\cdot) = P(T_2 > t|\cdot)$. The survival function is in the form

$$S(t_1, t_2|\lambda, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = e^{-\lambda t_3 - j} S_j(t_j - t_{3-j}|\boldsymbol{\theta}_j) \mathbf{1}_{\{0 \leq t_{3-j} \leq t_j\}}, \quad (1.6)$$

This implies that (2.a) The minimal lifetime T is independent of the residual lifetime $|T_1 - T_2|$ and the indicator $\mathbf{1}_{\{T_1 < T_2\}}$; and (2.b) T is exponentially distributed.

One immediate example is the well-cited Marshall-Olkin (1967) BVE model which satisfies both MLMP and JLMP. However, as shown by Block and Basu (1974), without the independence between T_1 and T_2 , absolute continuity (i.e., joint pdf of (T_1, T_2) exists and $P(T_1 = T_2) = 0$), MLMP and JLMP cannot exist simultaneously. More exactly, under dependence, if any of two among MLMP, JLMP, and absolute continuity are satisfied, the joint density must be lack of the other property. Marshall-Olkin's BVE retains the dependence instead. In contrast, the following BVE models ensure the absolute continuity by giving up either MLMP or JLMP.

Freund's (1961) absolute continuous bivariate exponential model (ACBVE) was originally designated for the parallel system. Let the model be characterized by parameters $(\alpha, \alpha', \beta, \beta')$, where $0 < \alpha < \alpha' < \alpha + \beta$, $0 < \beta < \beta' < \alpha + \beta$. It assumes that T_1 and T_2 follow independent exponential distributions with constant failure rates, α and β , respectively, when the system is under normal process. The dependence initiates when any one of the two components collapses. More specifically, the failure rate of component 1 increases to α' right after component 2 fails. Similarly,

failure rate of component 2 increases by $(\beta' - \beta)$ while component 1 fails. This phenomenon exhibits in the joint probability density

$$f(t_1, t_2 | \alpha, \alpha', \beta, \beta') = \begin{cases} \alpha\beta' e^{-(\alpha+\beta-\beta')t_1-\beta't_2} & 0 < t_1 \leq t_2 \\ \beta\alpha' e^{-(\beta+\alpha-\alpha')t_2-\alpha't_1} & 0 < t_2 \leq t_1 \end{cases}. \quad (1.7)$$

Rephrasing hazard rate by workload (in unit time), Freund's BVE explains the system with constant workload $(\alpha + \beta)$, in which the load is dynamically transferred between two components. When one of the components fails, the load for this component is reduced (by the amount of $(\beta' - \beta)$ for component 1 and $(\alpha' - \alpha)$ for component 2). The excess load $(\beta' - \beta)$ (or $(\alpha' - \alpha)$) then shifts to the partner component. Then

$$S(t_1, t_2 | \alpha, \alpha', \beta, \beta') = e^{-(\alpha+\beta)t_{3-j}} S_j(t_j - t_{3-j} | \alpha, \alpha', \beta, \beta') \mathbf{1}_{[0 < t_{3-j} \leq t_j]}, \quad (1.8)$$

where

$$\begin{aligned} S_1(t | \alpha, \alpha', \beta, \beta') &= \frac{\beta}{\alpha + \beta - \alpha'} e^{-\alpha't} + \frac{(\alpha - \alpha')}{\alpha + \beta - \alpha'} e^{-(\alpha+\beta)t}, \\ S_2(t | \alpha, \alpha', \beta, \beta') &= \frac{\alpha}{\alpha + \beta - \beta'} e^{-\beta't} + \frac{(\beta - \beta')}{\alpha + \beta - \beta'} e^{-(\alpha+\beta)t}. \end{aligned}$$

Apparently, JLMP is confirmed by (1.5). The marginals of T_1 and T_2 are no longer exponentially distributed as expected. They are mixtures of two exponentials.

To achieve JLMP and nonsingularity, Block and Basu (1974) modified Marshall-Olkin's BVE by giving up MLMP. Their ACBVE has the survival probability given by

$$S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}) = e^{-(\lambda_1+\lambda_2+\lambda_{12})t_{3-j}} S_j(t_j - t_{3-j}) \mathbf{1}_{[0 < t_{3-j} \leq t_j]}, \quad (1.9)$$

where

$$\begin{aligned} S_1(t | \lambda_1, \lambda_2, \lambda_{12}) &= \frac{\lambda_1 + \lambda_2 + \lambda_{12}}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_{12})t} - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2+\lambda_{12})t}, \\ S_2(t | \lambda_1, \lambda_2, \lambda_{12}) &= \frac{\lambda_1 + \lambda_2 + \lambda_{12}}{\lambda_1 + \lambda_2} e^{-(\lambda_2+\lambda_{12})t} - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2+\lambda_{12})t}. \end{aligned}$$

It indeed is a restricted version of Freund's ACBVE. The constraints of parameters are: $\alpha = \lambda_1 + \frac{\lambda_1}{\lambda_1 + \lambda_2} \lambda_{12}$, $\beta = \lambda_2 + \frac{\lambda_2}{\lambda_1 + \lambda_2} \lambda_{12}$, $\alpha' = \lambda_1 + \lambda_{12}$, and $\beta' = \lambda_2 + \lambda_{12}$. Physically, the model assumes that the proportions of excess workloads for both components are equal, i.e., $\frac{\alpha' - \alpha}{\beta} = \frac{\beta' - \beta}{\alpha}$. Moreover, by expressing the survival function of the Marshall-Olkin BVE in Friday's (1977) form, one finds that the Block-Basu ACBVE is the absolutely continuous component in the expression.

Sarkar's (1987) version of ACBVE secures MLMP and absolute continuity. This, however, is no longer JLMP, nor the model has a clearly physical interpretation. Sarkar's joint survival function is given by

$$S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}) = e^{-(\lambda_2 + \lambda_{12})t_3 - j} \{1 - [1 - e^{-\lambda_j t_3 - j}]^{-r} [1 - e^{-\lambda_j t_j}]^{1+r}\} \mathbf{1}_{[0 < t_j \leq t_3 - j]},$$

where $r = \frac{\lambda_{12}}{\lambda_1 + \lambda_2}$.

1.3.1.2 Shock Models

The idea of "shocks" was developed based on the relationship between Poisson process and the exponential distribution (cf. Marshall and Olkin (1967) and Arnold (1975)). Imagine that the occurrence of each mortality force as random shocks. Shocks occur in time according to Poisson process. A failure is induced when the accumulated damage caused by shocks exceeds the threshold. Generally, two types of shocks are considered: fatal (noncumulative) shocks and cumulative shocks. These are distinguished by recovered (renewal) probability, denoted by $r(t|j)$ which measures the survival probability at time t given experience of j shocks. Let $N(t)$ denote the stochastic process that counts the occurrences of shocks. Thus, $r(t|j) = P(T > t | N(t-) = j)$, for $0 < t < \infty$, $0 \leq j \leq \infty$. For fatal shocks, $r(t|j) = 0$, for $0 < t < \infty$, $1 \leq j < \infty$. This implies that $N(t)$ achieves its maximum value at 1. On the other hand, $r(t|j) > 0$, for $0 < t < \infty$, $1 \leq j < \infty$, if the shocks are cumulative. The joint survival function for a shock model is determined by recovered

probabilities and the corresponding stochastic processes (or the distribution of time interval of successive shocks). We shall now make this concept more concrete through the following examples.

The derivation of the Marshall-Olkin (1967) BVE follows the fatal shock principle. Consider in the two-component series system, components are damaged by “fatal shocks” from different sources. Let $N_1(t)$, $N_2(t)$, and $N_{12}(t)$ respectively characterize the occurrence of shocks experienced by component 1 only, component 2 only, and both jointly. Suppose that $N_1(t)$, $N_2(t)$, and $N_{12}(t)$ are independent Poisson processes with respective intensities λ_1 , λ_2 , and λ_{12} . The term, intensity, stands for the frequency of occurrences of the undergoing stochastic process, or is equivalent to the hazard rate in the corresponding survival function. Denote the renewal probabilities associated with N_1 , N_2 , and N_{12} by r_1 , r_2 , and r_{12} , respectively. Since $r_1(t|j) = r_2(t|j) = r_{12}(t|j) = 0$ for $j \geq 1$, the joint survival function $S(t_1, t_2|.)$ is found from

$$\begin{aligned} & S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}) \\ &= P(T_1 > t_1, T_2 > t_2 | \lambda_1, \lambda_2, \lambda_{12}) \\ &= P(N_1(t_1) = 0, N_2(t_2) = 0, N_{12}(\max\{t_1, t_2\}) = 0 | \lambda_1, \lambda_2, \lambda_{12}). \end{aligned}$$

Under the assumption of independence among N_1 , N_2 , and N_{12} , one obtains

$$S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)\}. \quad (1.10)$$

Thus, MLMP can be quickly identified. Rewrite (1.10) in the form of (1.6), then

$$S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}) = e^{-(\lambda_1 + \lambda_2 + \lambda_{12})t_{3-j}} e^{-(\lambda_j + \lambda_{12})(t_j - t_{3-j})} \mathbf{1}_{[0 < t_{3-j} \leq t_j]}. \quad (1.11)$$

One can further validate its JLMP.

“Cumulative shocks” model is exemplified by Downton’s (1970) BVE. This shock model comprises of two independent Poisson shocks $N_1(t)$ and $N_2(t)$, of which the

joint distribution of the occurrence of shocks follows a bivariate geometric structure¹ such that $r_1(t|j) = r_2(t|j) = \eta$, for $0 < t < \infty$, $0 \leq j < \infty$. Under the constant recovering probability assumption, exponential marginals are acquired. The derived joint lifetime density is given by

$$f(t_1, t_2 | \lambda_1, \lambda_2, \eta) = \frac{1}{(1-\eta)\lambda_1\lambda_2} \exp\left[-\frac{\lambda_1 t_1 + \lambda_2 t_2}{(1-\eta)}\right] B_0\left(\frac{2(\eta\lambda_1\lambda_2 t_1 t_2)^{\frac{1}{2}}}{1-\eta}\right), \quad (1.12)$$

where B_n is the modified Bessel function² of the first kind of order n .

The “weighted shocks model” is attributed to Ryu (1993). This development features a class of bivariate lifetime models with marginally progressing hazard rates. Let $N_1(t)$, $N_2(t)$, and $N_{12}(t)$ be defined as earlier. In the generalized version, Ryu modeled the hazard rates as weighted sum of cumulative shocks, say $d_1 N_1(t) + s_1 N_{12}(t)$ for component 1 and $d_2 N_2(t) + s_2 N_{12}(t)$ for component 2. It is the weights that give the flexibility for modeling. These weights can be analogous to the recovery probabilities, i.e., if the weight is finite, the associated shock is cumulative, otherwise it is noncumulative. In general, the survival function is given by

$$\begin{aligned} & S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}) \\ &= \exp\{-(\lambda_j + \lambda_{12})t_j - \lambda_j t_{3-j} + \frac{\lambda_j}{d_j}[1 - e^{-d_j t_j}] + \frac{\lambda_{3-j}}{d_{3-j}}[1 - e^{-d_{3-j} t_{3-j}}]\} \\ &\times \exp\left\{\frac{\lambda_{12}}{s_j}[1 - e^{-s_j(t_j - t_{3-j})}] + \frac{\lambda_{12}}{(s_1 + s_2)}[e^{-s_j(t_j - t_{3-j})} - e^{-s_{3-j} t_{3-j} - s_j t_j}]\right\} \mathbf{1}_{[0 < t_{3-j} \leq t_j]}. \end{aligned}$$

Thus, there is no indication of JLMP from this probability presentation. Nevertheless, it enjoys both JLMP and MLMP to a certain degree of approximation. More specifically, the Marshall-Olkin’s BVE is found at the boundary of this class, where $s_1 = s_2 = d_1 = d_2 = \infty$.

¹It implies that marginally, the total number of shocks that bring failure follows the geometric distribution.

² $B_n(x) = \sum_{r=0}^{\infty} \frac{(x/2)^{n+2r}}{r!(n+r)!}$.

1.3.1.3 Gumbel's Models

Gumbel (1960) proposed two types of BVE models purely from the considerations of exponential marginals and certain relationship between the two variates. Thus, only MLMP is clear in this motivation. We shall name these two BVE's as Gumbel-A BVE and Gumbel-B BVE. The respective joint survival functions are given by

$$S_{G_A}(t_1, t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} t_1 t_2}, \quad (1.13)$$

$$S_{G_B}(t_1, t_2) = e^{-[(\lambda_1 t_1)^{1/r} + (\lambda_2 t_2)^{1/r}]^r}. \quad (1.14)$$

It turns out that both Gumbel BVE's are ACBVE with MLMP. The Gumbel-B BVE satisfies (2.a), yet it does not possess the JLMP.

1.3.2 Bivariate Weibull Distributions

A typical technique used for constructing a bivariate Weibull distribution (BVW) is to make power transformation of marginals of BVE. Other alternatives, for instance, random hazard modeling and function-specific modeling are also commonly used. These BVW's will be introduced next.

1.3.2.1 Power Transformation

Lee (1979) made the transformation from the Gumbel-B BVE. The Marshall-Olkin BVE was generalized to a BVW with joint survival function

$$S(t_1, t_2 | \lambda_1, \lambda_2, \lambda_{12}, r_1, r_2) = \exp\{-\lambda_1 t_1^{r_1} - \lambda_2 t_2^{r_2} - \lambda_{12} \max(t_1^{r_1}, t_2^{r_2})\}. \quad (1.15)$$

By restricting $r_1 = r_2 = r$ in (1.15), Lee and Thompson's (1974) BVW is obtained.

1.3.2.2 Random Hazard Models

BVW can also be generated from random hazard models subject to Weibull marginals. The underlying hazards are adjusted multiplicatively by the random environmental hazard w . Essentially, this modeling strategy is similar to that of the frailty model. In Vaupel's (1979) frailty model, two lifetimes are conditionally independent of each other. The random hazard model instead allows conditional dependence. Let H_1 and H_2 be the respective cumulative hazard functions for T_1 and T_2 , and r ($0 < r \leq 1$) be the shape parameter. Given the random quantity w , the conditional survival function has the form

$$S(t_1, t_2 | w, r, \theta) = \exp \left\{ - \{ [H_1(t_1 | \theta)]^{1/r} + [H_2(t_2 | \theta)]^{1/r} \}^r w \right\}.$$

Let $\pi(w)$ be the pdf for w . Then the marginal survival function, or the Laplace transformation of w evaluated at $\{ [H_1(t_1 | \theta)]^{1/r} + [H_2(t_2 | \theta)]^{1/r} \}^r$, is given by

$$S(t_1, t_2 | r, \theta) = \int \exp \left\{ - \{ [H_1(t_1 | \theta)]^{1/r} + [H_2(t_2 | \theta)]^{1/r} \}^r w \right\} \pi(w) dw.$$

First assume that $w \sim G(a, 1)^3$ and Weibull marginals for $T_1 \sim W(\theta_1, \lambda_1)^4$ and $T_2 \sim W(\theta_2, \lambda_2)$, the joint survival function is derived as

$$S(t_1, t_2) = \{ 1 + [(e^{(\lambda_1 t_1)^{\theta_1}} - 1)^{1/r} + (e^{(\lambda_2 t_2)^{\theta_2}} - 1)^{1/r}]^r \}^{-a}. \quad (1.16)$$

This survival function was considered by Lu and Bhattacharyya (1990). Oakes (1982) has discussions for the special case when $r = 1$.

Next, assume that w follows a positive stable density with a scale parameter a , i.e., $E[\exp(-wy)] = \exp(-ay)$. Marginal distributions of T_1 and T_2 remain the same as above. The joint survival function becomes

$$S(t_1, t_2) = \exp \{ - [(\lambda_1 t_1)^{\theta_1/ar} + (\lambda_2 t_2)^{\theta_2/ar}]^{ar} \}, \quad (1.17)$$

³ $G(a, b)$ denotes gamma distribution with shape parameter a and scale parameter b , ($a, b > 0$).

⁴ $W(a, b)$ denotes Weibull distribution with shape parameter a and scale parameter b , ($a, b > 0$).

provided that $(a, r) \in (0, 1] \times (0, 1]$. (1.17) can also be obtained by making power transformation from the Gumbel-B ACBVE model (see (1.14)). Hougaard's (1986) derivation is a special case of (1.17), where $r = 1$.

1.3.2.3 Function-Specified Models

Lu and Bhattacharyya (1990) generalized two classes of absolute continuous BVW models, of which the joint survival functions are given by the expression

$$S(t_1, t_2) = \exp\{-(\lambda_1 t_1)^{\theta_1} - (\lambda_2 t_2)^{\theta_2} - \theta_0 g(t_1, t_2)\}, \quad (1.18)$$

where $g(t_1, t_2)$ is differentiable and could depend on other parameters.

Case 1: They first considered $\theta_0 \geq 0$ and $g(t_1, t_2) = [(\lambda_1 t_1)^{\theta_1/m} + (\lambda_2 t_2)^{\theta_2/m}]^m$ ($0 < m \leq 1$) in (1.18). Then

$$S(t_1, t_2) = \exp\{-(\lambda_1 t_1)^{\theta_1} - (\lambda_2 t_2)^{\theta_2} - \theta_0 [(\lambda_1 t_1)^{\theta_1/m} + (\lambda_2 t_2)^{\theta_2/m}]^m\}. \quad (1.19)$$

Case 2: Before introducing case 2, two dependent structures are defined. Random variables X and Y are said to be *positive quadrant dependence* PQD (Lehman, 1966) if for all (x, y) ,

$$P(X > x, Y > y) \geq P(X > x)P(Y > y). \quad (1.20)$$

Negative quadrant dependence (NQD) is correspondingly defined by reversing the inequality in (1.19). Most of the BVW's with survival function in the form of $\exp\{k(x, y)\}$ are either strictly PQD or NQD. For example, the Gumbel-A ACBVE is strictly NQD, while the BVW given in (1.17) is strictly PQD. A bivariate lifetime model that admits both PQD and NQD is pursued with the joint survival function

$$S(t_1, t_2) = S_1(t_1)S_2(t_2)e^{\theta_0\{1-S_1(t_1)\}\{1-S_2(t_2)\}}. \quad (1.21)$$

A NQD survival function is obtained under $-1 \leq \theta_0 < 0$, while if $\theta_0 > 0$, it is a PQD survival function. Particularly, let S_1, S_2 be the Weibull survival functions. A BVW

is then constructed with joint survival function

$$S(t_1, t_2 | \lambda_1, \lambda_2, \theta_0, \theta_1, \theta_2) = \exp\{-(\lambda_1 t_1)^{\theta_1} - (\lambda_2 t_2)^{\theta_2} + \theta_0[1 - e^{-(\lambda_1 t_1)^{\theta_1}}][1 - e^{-(\lambda_2 t_2)^{\theta_2}}]\}.$$

1.3.3 Use of Concomitant Information

To account for the presence of concomitant information in lifetime analysis, we briefly review some typical strategies considered in the past. In the classic regression modeling for one-dimensional response, the expectation of the response variable conditional on knowledge of the covariates (say vector $\mathbf{z}$) is assumed to be some function (usually linear) of the covariates. Feigl and Zelen (1965) and Zippin and Armitage (1966) recommended a similar adjustment for the mean lifetime by equating

$$E(T|\mathbf{z}) = \mathbf{z}'\boldsymbol{\beta}.$$

However, since lifetime is always positive, the range of $\boldsymbol{\beta}$ is restricted. As a remedy, Cox (1972) proposed a general proportional hazard (or accelerated lifetime) model via adjusting baseline hazard rate with a positive multiplier, a function of covariates, say $g(\mathbf{z}, \boldsymbol{\beta})$. The pdf of the survival time is taken as

$$p(t|\mathbf{z}) = h_0(t)g(\mathbf{z}, \boldsymbol{\beta}) \exp\{-H_0(t)g(\mathbf{z}, \boldsymbol{\beta})\}.$$

With $g(\mathbf{z}, \boldsymbol{\beta}) = e^{\mathbf{z}'\boldsymbol{\beta}}$, the range of $\boldsymbol{\beta}$ is now free from the covariates $\mathbf{z}$. Klein and Basu (1982) adopted the proportional hazard structure for independent Weibull variates.

The contemporary GLM technique will be described in detailed in Chapter 3. For example, Wada, Sen, and Simakura (1996) employed Sarkar's ACBVE model, and adjusted parameters with covariates $\mathbf{z}$ via logit regression model and proportional hazard model. These models are given by

$$\lambda_1(\mathbf{z})/\lambda_2(\mathbf{z}) = \exp(\mathbf{z}'\boldsymbol{\gamma}), \quad \lambda(\mathbf{z}) = \lambda_1(\mathbf{z}) + \lambda_2(\mathbf{z}) + \lambda_{12}(\mathbf{z}) = \exp(\mathbf{z}'\boldsymbol{\beta}). \quad (1.22)$$

1.3.4 Statistical Inference

Inference for competing risks models has mostly been classical. The maximum likelihood estimators (MLE) and the corresponding estimated asymptotic variance-covariance matrices were often used for the construction of tests and confidence intervals. The estimated asymptotic variance-covariance matrix is the inverse of the Fisher information matrix evaluated at the MLE's of the parameters. However, this approach demands the use of profile likelihood or the integrated likelihood when the full likelihood is not fully identifiable. This approach can be challenged by the non-uniqueness of MLE's and the difficulty of asymptotics. We illustrate below for some simple competing risks models to see how such estimators are obtained. Let $T_j = \min\{T_{j,1}, T_{j,2}\}$, and (t_j, i_j) denotes the realization of (T_j, I_j) , for $1 \leq j \leq n$. Denote $n_1 = \sum_{j=1}^n \mathbf{1}_{[t_{j,1} < t_{j,2}]}$, $n_2 = \sum_{j=1}^n \mathbf{1}_{[t_{j,2} < t_{j,1}]}$, and $n_{12} = \sum_{j=1}^n \mathbf{1}_{[t_{j,1} = t_{j,2}]}$. Notice that $n_{12} = 0$ under the absolutely continuous assumption.

Example 1 We begin the first example with the Marshall-Olkin BVE model given in (1.10). The joint likelihood function is given by

$$L(\lambda_1, \lambda_2, \lambda_{12}) \propto \lambda_1^{n_1} \lambda_2^{n_2} \lambda_{12}^{n_{12}} \exp\left\{-\lambda \sum_{j=1}^n t_j\right\}, \quad (1.23)$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$. The MLE's are

$$\hat{\lambda}_1 = n_1 \left(\sum_{j=1}^n t_j\right)^{-1}, \quad \hat{\lambda}_2 = n_2 \left(\sum_{j=1}^n t_j\right)^{-1}, \quad \hat{\lambda}_{12} = n_{12} \left(\sum_{j=1}^n t_j\right)^{-1}.$$

The estimated asymptotic variance-covariance matrix follows

$$\mathbf{I}^{-1}(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_{12}) = \begin{bmatrix} \hat{\lambda}_1 & 0 & 0 \\ 0 & \hat{\lambda}_2 & 0 \\ 0 & 0 & \hat{\lambda}_{12} \end{bmatrix}.$$

In this example, MLE's and their asymptotic behaviors can be easily verified.

Example 2 Consider the Gumbel-B ACBVE model given in (1.14). The joint likelihood function is given by

$$L(\lambda, \rho, r) \propto \lambda^n \rho^{n_1} (1 - \rho)^{n_2} \exp\left\{-\lambda \sum_{j=1}^n t_j\right\} \mathbf{1}_{[0 < r \leq 1]}, \quad (1.24)$$

where $\lambda = (\lambda_1^{1/r} + \lambda_2^{1/r})^r$ and $\rho = \frac{\lambda_1^{1/r}}{(\lambda_1^{1/r} + \lambda_2^{1/r})}$. Clearly, this likelihood is nonidentifiable.

As a consequence, the Fisher information matrix $\mathbf{I}(\lambda_1, \lambda_2, r)$ is singular. The MLE's are

$$\hat{\lambda} = n \left(\sum_{j=1}^n t_j \right)^{-1}, \quad \hat{\rho} = n_1/n. \quad (1.25)$$

According to the solution, there is no unique MLE for r , and the asymptotic theory can not be applied. Suppose that r can be treated as a nuisance parameter. One may approach this problem from either the profile likelihood or the integrated likelihood angle. Both shall also provide the same MLE's as above. However, r may carry an important message for the underlying study. Then how does one provide a solution?

Example 3 Let assume Block-Basu's ACBVE model given in (1.10). The likelihood function for $\{(t_j, i_j) : 1 \leq j \leq n\}$ is given by

$$L(\lambda, \rho, \phi) \propto \lambda^n \rho^{n_1} (1 - \rho)^{n_2} \exp\left\{-\lambda \sum_{j=1}^n t_j\right\} \mathbf{1}_{[0 < \phi < \lambda]}, \quad (1.26)$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$, $\phi = \lambda_1 + \lambda_2$, and $\rho = \frac{\lambda_1}{(\lambda_1 + \lambda_2)}$. For a fixed ϕ , the MLE's are

$$\hat{\lambda} = \min\left\{\phi, n\left(\sum_{j=1}^n t_j\right)^{-1}\right\}, \quad (1.27)$$

$$\hat{\rho} = n_1/n. \quad (1.28)$$

The Fisher information matrix can not be obtained since $\log(\mathbf{1}_{[0 < \phi < \lambda]})$ is not differentiable with respect to ϕ or λ . The same problem which we encountered with in Example 2 arises again.

Example 4 Consider the covariate-adjusted Sarkar BVE model. Wada, Sen, and Simakura's (1996) ignored the parameter which is nonidentifiable in the likelihood function, and approached directly from the integrated likelihood function. MLE's are obtained from the logarithm of the integrated likelihood function

$$\sum_{j=1}^n \{ \mathbf{z}'_j \boldsymbol{\beta} - e^{\mathbf{z}'_j \boldsymbol{\beta}} t_j + \mathbf{1}_{\{t_j=1\}} \mathbf{z}'_j \boldsymbol{\gamma} - \log(1 + e^{\mathbf{z}'_j \boldsymbol{\gamma}}) \}. \quad (1.29)$$

Since no closed form solution for the MLE's of the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ is available, one employs the iterative method, Newton-Raphson approach. Wada and Sen (1993) verified the consistency of the MLE's, i.e., $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ has asymptotic multivariate normal distribution with mean vector zero and variance-covariance matrix $[\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\gamma})]^{-1}$.

In contrast, a Bayesian utilizing both sample information and non-sample information makes inference on parameters from the conditional perspective. The non-sample information is given by the parameter distribution functions, namely, prior $\pi(\cdot)$ and loss functions $L_s(\cdot)$ (we shall assume square error loss throughout this study). To demonstrate the idea, let $\mathbf{d}^*$ and $\boldsymbol{\theta}$ denote the data and model parameters, respectively. The Bayesian inference is based on the (proper) posterior distribution, i.e., $\pi(\boldsymbol{\theta}|\mathbf{d}^*) \propto L(\boldsymbol{\theta}|\mathbf{d}^*) \pi(\boldsymbol{\theta})$. The Bayes estimator $\hat{\boldsymbol{\theta}}_B$ for $\boldsymbol{\theta}$ is calculated by minimizing the posterior Bayes risk $E[L_s(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_B|\mathbf{d}^*)]$, equivalently, $\hat{\boldsymbol{\theta}}_B = E_{\pi}(\boldsymbol{\theta}|\mathbf{d}^*)$. The corresponding posterior variance is $V(\boldsymbol{\theta}|\mathbf{d}^*) = E_{\pi}(\hat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}|\mathbf{d}^*)^2$. The nonidentifiability problem can be circumvented via a suitable prior specification. How could $\pi(\boldsymbol{\theta})$ be determined for competing risks models? We review some of the available literature.

Liu and Nakao (1991) suggested the gamma-related prior for the Marshall-Olkin BVE, i.e.,

$$\pi(\lambda_1, \lambda_2, \lambda_{12}) \propto \lambda_1^{a_1} (\lambda_{12} + \lambda_2)^{a_2} \lambda_2^{a_2} (\lambda_{12} + \lambda_1)^{a_2} \lambda_{12}^{a_3} e^{-a_4 \lambda_1 - a_5 \lambda_2 - a_6 \lambda_{12}}. \quad (1.30)$$

The derived Bayes estimators for $(\lambda_1, \lambda_1, \lambda_{12})$ and $S(t_1, t_2 | \lambda_1, \lambda_1, \lambda_{12})$ can all be expressed explicitly. Ideally, information for $(a_1, a_1, a_2, a_3, a_4, a_5, a_6)$ may be historically available. This does not always happen in practice.

Lu (1992) proposed the data-driven prior for the Marshall-Olkin BVW model (1.15) with equal shape parameter r . He considered a two-stage prior given by

$$\begin{aligned} \pi_{\mathbf{b}}(r = b_j) &= p_j \\ \pi_{\lambda_j}(\lambda_1, \lambda_2, \lambda_{12} | r = b_j) &\propto (\lambda_{12} + \lambda_1)^{a_{1j}} (\lambda_{12} + \lambda_2)^{a_{2j}} e^{-\xi_{1j} \lambda_1 - \xi_{2j} \lambda_2 - (\xi_{1j} + \xi_{2j}) \lambda_{12}}, \end{aligned}$$

where $\{b_j, p_j\}_{j=1}^2$ and $\{a_{ij}, \xi_{ij}\}_{(i,j) \in \{1,2\}}$ are estimated from the data. Explicit posterior distributions are obtained. Other subjective options, such as the gamma prior (Berger and Sun, 1993) and the joint location-scale conjugate prior (Bury, 1973) have been used. These do not provide a closed form posterior distribution. However, the computation can be accomplished by numerical methods, such as Gibbs sampler and adaptive-rejection sampler.

1.4 Noninformative Priors

In view of the difficulty of finding subjective priors, one has to restrict attention to noninformative priors. The simplest and the most historical noninformative prior is attributed to Laplace (1820). He vaguely assigns equal probability on the entire parameter space, i.e., $\pi_U(\boldsymbol{\theta}) \propto 1$. However, a constant prior for one parameterization will not typically transform into a constant for another. Thus, Jeffreys (1961) proposed a prior which is invariant under any one-to-one reparameterization. It is derived such that it is proportional to the positive square root of the determinant of the

Fisher information matrix. Let $\{Y_i\}$ be independently distributed random variables with identical pdf $f(\cdot|\boldsymbol{\theta})$. Denote $l(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \log[f(\mathbf{y}_i|\boldsymbol{\theta})]$, then

$$\pi_J(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}, \quad (1.31)$$

where $\mathbf{I}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[-\partial^2 l(\boldsymbol{\theta})/\partial\theta_u\partial\theta_v]$.

Subsequent derivation for noninformative priors are mostly for pragmatic purposes. For instance, interest lies in particular parameters, or sometimes in functions of parameters. Bernardo (1979) introduced the notion of the reference priors. He recommended $Q = E\{\log[\pi(\boldsymbol{\theta}|\mathbf{y})] - \log[\pi(\boldsymbol{\theta})]\}$ as a measure of information about $\boldsymbol{\theta}$ provided by the data. Here E denotes expectation over the joint distribution of $\mathbf{y}$ and $\boldsymbol{\theta}$. The rationale behind is that the larger this quantity Q , the less informative the prior. The reference prior $\pi_R(\boldsymbol{\theta})$ is the prior that maximizes Q . To construct a reference prior, one could follow the *four-step* scheme proposed by Berger and Bernardo (1989).

Step.1 Let $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{[-1]})$ be a partition of the parameter, where $\boldsymbol{\theta}_1$ are the parameters of interest, and $\boldsymbol{\theta}_{[-1]}$ are the nuisance parameters. First, the conditional prior, $\pi(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)$, is defined as

$$p_1(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1) = |\mathbf{I}_{22}(\boldsymbol{\theta})|^{1/2}, \quad (1.32)$$

where the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ is partitioned as

$$\mathbf{I} \equiv \mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_{11}(\boldsymbol{\theta}) & \mathbf{I}_{12}(\boldsymbol{\theta}) \\ \mathbf{I}_{21}(\boldsymbol{\theta}) & \mathbf{I}_{22}(\boldsymbol{\theta}) \end{bmatrix}.$$

Step.2 Choose a sequence of subsets of the parameter space Ω for $\boldsymbol{\theta}$, $\Omega_1 \subset \Omega_2 \subset \dots$, such that $\cup_i \Omega_i = \Omega$, and $p_1(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)$ has finite mass on $\Omega_{\mathbf{i}}\boldsymbol{\theta}_1 = \{\boldsymbol{\theta}_{[-1]} : (\boldsymbol{\theta}_1, \boldsymbol{\theta}_{[-1]}) \in \Omega_i\}$ for all $\boldsymbol{\theta}_1$. Normalize $p_1(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)$ on each $\Omega_{\mathbf{i}}\boldsymbol{\theta}_1$ to obtain

$$p_2^{(\mathbf{i})}(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1) = k_{\mathbf{i}}(\boldsymbol{\theta}_1)p_1(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)\mathbf{1}_{[\boldsymbol{\theta}_{[-1]} \in \Omega_{\mathbf{i}}\boldsymbol{\theta}_1]}, \quad (1.33)$$

where $k_i(\boldsymbol{\theta}_1) = \{\int_{\Omega_{i,\boldsymbol{\theta}_1}} p_1(\boldsymbol{\theta}_{|-1}|\boldsymbol{\theta}_1) d\boldsymbol{\theta}_{|-1}\}^{-1}$.

Step.3 Find the marginal reference prior (assuming the existence) for $\boldsymbol{\theta}_1$ with respect to $p_2^{(i)}(\boldsymbol{\theta}_{|-1}|\boldsymbol{\theta}_1)$. That is

$$p_3^{(i)}(\boldsymbol{\theta}_1) = \exp \left\{ \frac{1}{2} \int_{\Omega_{i,\boldsymbol{\theta}_1}} p_2^{(i)}(\boldsymbol{\theta}_{|-1}|\boldsymbol{\theta}_1) \log \left[\frac{\mathbf{I}(\boldsymbol{\theta})}{\mathbf{I}_{22}(\boldsymbol{\theta})} \right] d\boldsymbol{\theta}_{|-1} \right\}. \quad (1.34)$$

Step.4 Define the reference prior for $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{|-1})$ by

$$\pi_R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{|-1}) = \lim_{i \rightarrow \infty} \left\{ \frac{k_i(\boldsymbol{\theta}_1) p_3^{(i)}(\boldsymbol{\theta}_1)}{k_i(\boldsymbol{\theta}_1^*) p_3^{(i)}(\boldsymbol{\theta}_1^*)} \right\} p_1(\boldsymbol{\theta}_{|-1}|\boldsymbol{\theta}_1), \quad (1.35)$$

assuming the limit exists and $\boldsymbol{\theta}_1^*$ is a fixed point.

Remark: When there is no nuisance parameter, i.e., $\boldsymbol{\theta}_{|-1}$ is an empty vector, (1.35) equals to Jeffreys' prior.

A somewhat different criterion for developing noninformative priors is based on matching the posterior coverage probability of a Bayesian credible set with the corresponding frequentist coverage probability with a reminder. The matching, discussed further, is accomplished through either posterior quantiles or the highest posterior density (HPD) ellipsoid. Matching based on one-dimensional posterior quantiles with order $o(n^{-1/2})$ was first investigated by Welch and Peers (1963). Let $\theta_1^{(\alpha)}(\pi, \mathbf{d}^*)$ denote the α^{th} posterior quantile for θ_1 . If a prior $\pi(\boldsymbol{\theta})$ satisfies $P\{\theta_1 \leq \theta_1^{(\alpha)}(\pi, \mathbf{d}^*) | \boldsymbol{\theta}\} = \alpha + o(n^{-1/2})$, it is called the first order probability matching prior, and is characterized by

$$\sum_{j=1}^k \frac{\partial}{\partial \theta_j} [(I^{11})^{-1/2} I^{11} \pi(\boldsymbol{\theta})] = 0, \quad (1.36)$$

where $\mathbf{I}^{-1} = ((I^{ij}))$. For $k = 1$, the unique solution $\pi(\boldsymbol{\theta})$ for (1.36) is Jeffreys' prior. Tibshirani (1989) characterized the class of the first order probability matching priors

when $I^{jj} = 0 \forall j = 2, \dots, k$. For $g(\theta_2, \dots, \theta_k) > 0$, the solution for (1.36) reduces to

$$\pi(\boldsymbol{\theta}) \propto I^{11}(\boldsymbol{\theta})g(\theta_2, \dots, \theta_k). \quad (1.37)$$

For better approximation, Dey and Mukerjee (1993) and Mukerjee and Ghosh (1997) developed the second order probability matching prior with remainder of order $o(n^{-1})$. This development helps narrow down the selection of priors, and quite often leads to a unique prior within the first order probability matching priors. In addition to (1.36), it is obtained by solving

$$\sum_{j=1}^k \sum_{r=1}^k \frac{\partial^2}{\partial \theta_j \partial \theta_r} \{ \tau^{jr} \pi(\boldsymbol{\theta}) \} - \frac{1}{3} \sum_{v=1}^k \sum_{j=1}^k \sum_{r=1}^k \sum_{u=1}^k \frac{\partial}{\partial \theta_v} \{ L_{jru} \tau^{jr} (3I^{uv} - 2\tau^{uv}) \pi(\boldsymbol{\theta}) \} = 0, \quad (1.38)$$

where $\tau^{jr} = I^{jj} I^{rr} / I^{11}$, and $L_{jru} = E_{\boldsymbol{\theta}} \{ \partial^3 l(\boldsymbol{\theta}) / \partial \theta_j \partial \theta_r \partial \theta_u \}$. We simplified (1.38) for a diagonal $\mathbf{I}(\boldsymbol{\theta})$ to

$$\frac{1}{6} g(\theta_2, \dots, \theta_k) \frac{\partial}{\partial \theta_1} [I_{11}^{-3/2} L_{1,1,1}] + \sum_{u=1}^k \frac{\partial}{\partial \theta_u} [I_{11}^{-1/2} I_{uu}^{-1} L_{11u} g(\theta_2, \dots, \theta_k)] = 0. \quad (1.39)$$

Further results in this area were pursued by Datta (1996). Datta (1996) derived necessary and sufficient conditions for a prior matching the marginal coverage probability (up to $O(n^{-1})$) for each parameter of interest to be jointly probability matching. He refers to such priors as jointly-probability-matching priors. In particular, if $\mathbf{I}(\boldsymbol{\theta})$ is diagonal, the conditions are automatically satisfied.

An alternative approach is to ensure the frequentist validity of the highest posterior density region (HPD), which provides a matching prior when the parameter is multi-dimensional. Ghosh and Mukerjee (1993) characterized the HPD prior, $\pi_{MH}(\boldsymbol{\theta})$ with matching up to $o(n^{-1})$, as a solution of the differential equation

$$\sum_{j=1}^k \sum_{r=1}^k \frac{\partial}{\partial \theta_j} \{ I^{jr} \pi_r(\boldsymbol{\theta}) \} + \sum_{v=1}^k \sum_{j=1}^k \sum_{r=1}^k \sum_{u=1}^k \frac{\partial}{\partial \theta_j} \{ I^{vr} I^{ju} L_{jru} \pi_{MH}(\boldsymbol{\theta}) \} = 0, \quad (1.40)$$

where $\pi_r(\boldsymbol{\theta}) = \partial\pi(\boldsymbol{\theta})/\partial\theta_r$ and $L_{j,ru} = E_{\boldsymbol{\theta}}[\{\partial l(\boldsymbol{\theta})/\partial\theta_j\}\{\partial^2 l(\boldsymbol{\theta})/\partial\theta_r\partial\theta_u\}]$, for $1 \leq j, r, u \leq k$. The simplification on (1.40) for a diagonal $\mathbf{I}(\boldsymbol{\theta})$ is then given by

$$\sum_{r=1}^k \frac{\partial}{\partial\theta_r} \{I^{rr}\pi_r(\boldsymbol{\theta})\} + \sum_{r=1}^k \sum_{u=1}^k \frac{\partial}{\partial\theta_r} \{I^{rr}I^{uu}L_{u,ru}\pi(\boldsymbol{\theta})\} = 0. \quad (1.41)$$

1.5 Research Proposal

To the extent of Bayesian method and model generalization in the competing risks framework, this study will focus on:

- Bayesian approach is often suitable in overcoming the nonidentifiability of the likelihood. Noninformative priors are often considered as acceptable substitutes when no informative prior is available. We shall also consider informative priors in Chapter 2 to resolve the nonidentifiability problem.
- The reparameterization invariant noninformative priors, such as Jeffreys' prior and the reference priors, are particularly preferable. Such priors, however, are derived based on a fully-specified and nonsingular Fisher information matrix. They cannot be directly applied to the competing risks data due to the singularity of the Fisher information matrix (see Example 2) or its inability to differentiate on the log-likelihood with respect to some of the parameters (see Example 3). To overcome this problem, we suggest the stagewise prior elicitation in Chapter 2. This technique is extended as well to the generalized BVE models introduced in Chapter 3, where the baseline model is adjusted for covariates. These stagewise priors will be further examined by the probability matching criteria for comparison purpose.
- Posterior distributions are the cornerstone of Bayesian inference. It is important that posteriors are genuine probability distributions. In both these chapters, we shall verify the propriety of the posteriors, and implement the Bayesian method for some

real life data using the Markov Chain Monte Carlo method. In addition, model adequacy will be tested through Bayesian posterior goodness-of-fit procedure.

- A more flexible class of models for the competing risks outcomes is suggested in Chapter 4. One of the advantages of the proposed methodology in this chapter is that it always provides an identifiable likelihood. Noninformative priors can, nevertheless, be used fruitfully.

CHAPTER 2

BAYESIAN ANALYSIS OF SELECTED BIVARIATE EXPONENTIAL MODELS

2.1 Introduction

Bivariate exponential (BVE) distributions were developed primarily for modeling correlated bivariate lifetimes. They have been used widely also for the analysis of competing risks data when the component risks are dependent. For such analysis, frequentist approach often runs into difficulty due to nonidentifiable likelihood. With an end to overcome the nonidentifiability, recent literature (cf. Section 1.3) has been geared towards Bayesian analysis with subjective priors. However, systematic prior elicitation is often difficult. This chapter focuses instead on Bayesian analysis with noninformative priors.

In our competing risks framework, we consider the subfamily of BVE models for which the time to failure T is independent of the associated the indicator variable I denoting the cause of failure. We will provide several examples of this in Section 2.2. In our framework, each single datum consists of three pieces of information: the time to failure T , the nonsingularity indicator Ψ denoting whether or not two components have simultaneously failed, and the cause-specific indicator Δ . Due to the nature of data and BVE model structure, quite often we are led a likelihood function with a nonregular Fisher information matrix impeding thereby the calculation of standard noninformative priors, such as Jeffreys' prior and its variants. This means that it is necessary to modify the priors. In Section 2.3, a stagewise noninformative prior elicitation strategy is proposed, and is utilized for a general class of nonsingular BVE models. A variety of noninformative priors are developed, and are used for

data analysis. In Sections 2.4 and 2.5, we demonstrate Bayesian applications for the Marshall-Olkin BVE and selected ACBVE's in detail. In addition, we compare the performance of newly developed noninformative priors in Section 2.6, particularly those based on the frequentist probability matching criterion. In Section 2.7, we provide a simulation study to demonstrate the versatility of Bayesian procedures.

2.2 Notation

The pair (T, I) actually reflects three facts: time to failure, cause of failure and singularity (or nonsingularity). For this reason, we define a triplet that is equivalent to (T, I) . Let $\Psi = 1$ if $I = 1$ exclusively or $I = 2$ exclusively, $\Psi = 0$ if $I = 1$ and $I = 2$ simultaneously, $\Delta = 1$ for $I = 1$ given $\Psi = 1$, and $\Delta = 0$ for $I = 2$ given $\Psi = 1$. Both Δ and Ψ have interesting interpretations in the current context: Δ is interpreted as the cause-specific indicator, and Ψ is the nonsingularity indicator. One should keep in mind that $P(\Psi = 1) = 1$ under the assumption of nonsingularity. Therefore, in this case, we shall not redundantly retain Ψ . Following the well-defined relationship between I and (Δ, Ψ) , the triplet presentation is given by (T, Δ, Ψ) . Here are some notions we shall use mostly in this study.

- n : sample size.
- (t_i, δ_i, ψ_i) : realization of (T_i, Δ_i, Ψ_i) , for $1 \leq i \leq n$.
- $\mathbf{d}_i \equiv (t_i, \delta_i, \psi_i)$: i^{th} observed triplet.
- $\mathbf{d} \equiv \{\mathbf{d}_i : 1 \leq i \leq n\}$: the entire dataset.
- t_t : realization of $T_t \equiv \sum_{i=1}^n T_i$.
- n_1 : realization of $N_1 \equiv \sum_{i=1}^n \Delta_i \Psi_i$.
- n_2 : realization of $N_2 \equiv \sum_{i=1}^n (1 - \Delta_i) \Psi_i$.
- n_{12} : realization of $N_{12} \equiv \sum_{i=1}^n (1 - \Psi_i)$.

- θ : model based parameter vector.
- θ_i : the i^{th} element of θ .
- $\mathbf{1}_{[E]}$: indicator function of event E .
- $f(x|\theta)$: probability density function (pdf) of x given parameters θ .
- $F(x|\theta)$: $\int_{-\infty}^x f(y|\theta) dy$.
- $L(\theta|\mathbf{d})$: likelihood function of θ given data $\mathbf{d}$.
- $P(\cdot)$: probability measure.
- $\pi(\theta)$: prior density of θ .
- $\pi(\theta|\mathbf{d})$: posterior density of θ given data $\mathbf{d}$ and prior π .
- $\oplus[\mathbf{A}_i]$: a block diagonal matrix with i^{th} diagonal element $\mathbf{A}_i$.

2.3 Noninformative Prior Modification

The Fisher information matrix has so far been the cornerstone in the development of a wide class of noninformative priors. However, for most competing risks models, the likelihood function of (T, Δ, Ψ) generated from ACBVE's becomes nonidentifiable, thereby making the direct calculation of the Fisher information matrix invalid. We begin with a class of likelihood functions. Assume that $\{\mathbf{D}_i^* : 1 \leq i \leq n\}$ are a sequence of independent random vectors from a common distribution with pdf $f(\cdot|\theta)$. Let $\mathbf{d}^* \equiv \{\mathbf{d}_i^* : 1 \leq i \leq n\}$ be the corresponding realization of $\{\mathbf{D}_i^* : 1 \leq i \leq n\}$. Write $\theta = (\theta_1, \theta_{[-1]})$ as a partition of θ . Suppose that the likelihood function $L(\theta|\mathbf{d}^*) = \prod_{i=1}^n f(\mathbf{d}_i^*|\theta)$ is given by the expression

$$L(\theta|\mathbf{d}^*) \equiv L_1(\theta_1|\mathbf{d}^*)L_0(\theta_{[-1]}|\mathbf{d}^*)\mathbf{1}_{[\theta_{[-1]} \in R(\theta_1)]}, \quad (2.1)$$

where $L_1(\cdot)$ and $L_0(\cdot)$ are some smooth differentiable functions, $R(\theta_1)$ defines the range for $\theta_{[-1]}$, which may depend on θ_1 . In case $L_0(\theta_{[-1]}|\mathbf{d}^*)$ is independent of $\theta_{[-1]}$,

and $\mathbf{1}_{\{\boldsymbol{\theta}_{[-1]} \in R(\boldsymbol{\theta}_1)\}}$ is present, the logarithm of (2.1) is nondifferentiable with respect to any element of $\boldsymbol{\theta}_{[-1]}$. Hence the Fisher information matrix does not exist. As a consequence, noninformative priors such as Jeffreys' prior, reference priors, or probability matching priors become impossible to calculate. This phenomenon as mentioned earlier is present for competing risks models originating out of ACBVE's, e.g., Freund's ACBVE and Block-Basu's ACBVE. To overcome this technical difficulty, we propose a two-stage noninformative prior elicitation scheme. This scheme stems from the idea of the prior decomposition (Berger and Bernardo, 1989), i.e., $\pi(\boldsymbol{\theta}) = \pi_0(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)$. The two-stage eliciting procedure is stated follows. Given $\boldsymbol{\theta}_1$, a "proper" noninformative prior $\pi_0(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)$ is assumed at the first stage. The marginal noninformative prior, $\pi_1(\boldsymbol{\theta}_1)$ say, is then derived from the marginal likelihood function

$$\int_{R(\boldsymbol{\theta}_1)} L(\boldsymbol{\theta}|\mathbf{d}^*)\pi_0(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1) d\boldsymbol{\theta}_{[-1]} \propto L_1(\boldsymbol{\theta}_1|\mathbf{d}^*). \quad (2.2)$$

The above holds trivially for any proper $\pi_0(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)$. Furthermore, the propriety of $\pi_0(\boldsymbol{\theta}_{[-1]}|\boldsymbol{\theta}_1)$ is essential to guarantee both the existence of $L_1(\boldsymbol{\theta}_1|\mathbf{d}^*)$ and the validity of the second stage procedure. In the second stage specification, using $L_1(\boldsymbol{\theta}_1|\mathbf{d}^*)$ as the base, we consider Laplace's prior, Jeffreys' prior, and the reference prior as default priors.

A general expression for the likelihood function based on a variety of BVE's is in the form of (2.1) and is given by

$$L(\lambda, \rho, \phi|\mathbf{d}) \propto \lambda^n e^{-\lambda t} \rho^{n_1} (1 - \rho)^{n_2} L_0(\phi|\mathbf{d}) \mathbf{1}_{\{\phi \in R(\lambda, \rho)\}}, \quad (2.3)$$

where λ is the hazard rate for T and ρ is the probability of failing from cause 1. For the Marshall-Olkin BVE, $L_0(\phi|\mathbf{d}) \propto \phi^{n_1+n_2} (1 - \phi)^{n_{12}}$, and $R(\lambda, \rho)$ is free from (λ, ρ) . For ACBVE's, $L_0(\phi|\mathbf{d}) = 1$ and $\mathbf{1}_{\{\phi \in R(\lambda, \rho)\}}$ is present. Noninformative prior calculation as introduced in Section 1.4 is possible for the Marshall-Olkin BVE. For ACBVE models, however, priors are found by following the two-stage scheme as described

above. We shall illustrate these two-stage priors for ACBVE models. First, however, we begin with the Marshall-Olkin BVE.

2.4 Bayesian Analysis for Marshall-Olkin BVE

Suppose that lifetimes of sampled objects are subject to two competing risks, where the chance of simultaneous failure is nonzero, i.e., $P(N_{12} > 0) > 0$. Suppose that the underlying model is the Marshall-Olkin BVE. Then the likelihood function under its original parameterization is given in (1.10). We consider the following reparameterization based on two primary objectives: the likelihood function has the expression compatible to (2.1), and the new parameters address certain model features. Here we choose

$$\lambda = \lambda_1 + \lambda_2 + \lambda_{12}, \quad \rho = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad \phi = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_{12}}.$$

These new parameters λ , ϕ , and ρ are interpreted as the instantaneous failure rate of T , the probability of failing from a single cause, and the conditional probability of failing from cause 1 given a single-cause failure, respectively. Given n_1 , n_2 , n_{12} , and t , the joint likelihood function of (λ, ρ, ϕ) is given by

$$L(\lambda, \rho, \phi | \mathbf{d}) \propto \lambda^n e^{-\lambda t} \rho^{n_1} (1 - \rho)^{n_2} \phi^{n_1+n_2} (1 - \phi)^{n_{12}}, \quad (2.4)$$

which corresponds to the likelihood function (2.1) by setting $\theta_1 = (\lambda, \rho)$, $\theta_{[-1]} = \phi$, $L_1(\theta_1 | \mathbf{d}) = L_1(\lambda, \rho | \mathbf{d}) \propto \lambda^n e^{-\lambda t} \rho^{n_1} (1 - \rho)^{n_2}$, and $L_0(\theta_{[-1]} | \mathbf{d}) = L_0(\phi | \mathbf{d}) \propto \phi^{n_1+n_2} (1 - \phi)^{n_{12}}$. On the other hand, by the independence of T and (Δ, Ψ) , one can write

$$L(\lambda, \rho, \phi | \mathbf{d}) = L_e(\lambda | \mathbf{d}) L_d(\rho, \phi | \mathbf{d}), \quad (2.5)$$

where $L_e(\lambda | \mathbf{d}) \propto \lambda^n e^{-\lambda t}$, and $L_d(\rho, \phi | \mathbf{d}) \propto \rho^{n_1} (1 - \rho)^{n_2} \phi^{n_1+n_2} (1 - \phi)^{n_{12}}$. This demonstrates the likelihood independence and consequent parameter orthogonality. We derive now certain noninformative priors based on (2.5).

2.4.1 Jeffreys' Prior and Probability Matching Property

Initially, the full Fisher information matrix is calculated as

$$\mathbf{I}(\lambda, \rho, \phi) = \begin{bmatrix} \mathbf{I}_c(\lambda) & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{2 \times 1} & \mathbf{I}_d(\rho, \phi) \end{bmatrix},$$

where $\mathbf{0}_{x,y}$ denotes the zero matrix of dimension $x \times y$, $\mathbf{I}_c(\lambda) = \lambda^{-2}$, and

$$\mathbf{I}_d(\rho, \phi) = \begin{bmatrix} \phi \rho^{-1} (1 - \rho)^{-1} & 0 \\ 0 & \phi^{-1} (1 - \phi)^{-1} \end{bmatrix}.$$

Thus, Jeffreys' prior is given by

$$\pi_J \equiv \pi_J(\lambda, \rho, \phi) \propto \pi_{J,c}(\lambda) \pi_{J,d}(\rho, \phi), \quad (2.6)$$

where

$$\pi_{J,c}(\lambda) \propto \lambda^{-1}, \quad (2.7)$$

$$\pi_{J,d}(\rho, \phi) \propto \rho^{-\frac{1}{2}} (1 - \rho)^{-\frac{1}{2}} (1 - \phi)^{-\frac{1}{2}}. \quad (2.8)$$

Following the arguments due to the orthogonality between λ and (ρ, ϕ) , Jeffreys' prior given in (2.6) is the unique second order probability matching prior for λ . In Section 2.7, we will have more extensive discussions on this matching property.

2.4.2 Reference Prior

The reference prior due to Bernardo (1979) and Berger and Bernardo (1989, 1992) is appropriate when a subset of (λ, ρ, ϕ) is the parameters of interest while the complementary subset consists of the nuisance parameters. The *four-step* formulae by Berger and Bernardo (1989) is used to compute the reference priors. Let $\Omega = \{(\lambda, \rho, \phi) : \lambda > 0, \rho \in (0, 1), \phi \in (0, 1)\}$. Select $\Omega_i = \{(i^{-1}, 1 - i^{-1}) \times (i^{-1}, 1 - i^{-1}) \times (i^{-1}, i)\}$ as the sequence of compact subspaces of Ω . Let θ_I and θ_N be generic symbols for

the parameters of interest and the nuisance parameters, respectively. The resulting reference priors, denoted by π_{Ref} , with respect to all possible nontrivial parameter vector partitions are presented in Table 2.1.

Table 2.1. Reference Priors for the Marshall-Olkin BVE

θ_I	λ	(ρ, ϕ)	ρ	(ϕ, λ)	ϕ	(λ, ρ)
θ_N	(ρ, ϕ)	λ	(ϕ, λ)	ρ	(λ, ρ)	ϕ
$\pi_{Ref}(\theta)$	π_J	π_J	$\pi_J \phi^{-\frac{1}{2}}$	$\pi_J \phi^{-\frac{1}{2}}$	$\pi_J \phi^{-\frac{1}{2}}$	$\pi_J \phi^{-\frac{1}{2}}$

Reference priors do not often agree with Jeffreys' prior. This can be found in columns 4-7 of Table 2.1. The coincidence occurs when the parameter vector is partitioned according to the variable type (discrete or continuous) that the parameter is associated with. Columns 2-3 in Table 2.1 exemplify this result.

2.4.3 Propriety of Posteriors and Bayesian Inference

It is necessary to check the propriety of posteriors under the above noninformative priors. Let $\pi_R \propto \pi_J \phi^{-\frac{1}{2}}$. Define the beta (proper or improper) pdf

$$\pi(x; a, b) \propto x^{a-1} (1-x)^{b-1}. \quad (2.9)$$

Then the noninformative priors π_J , π_R , and $\pi_U \equiv \pi_U(\lambda, \rho, \phi) \propto 1$ are all expressible in the form

$$\pi(\lambda; a_\lambda, 1) \pi(\rho; a_\rho, b_\rho) \pi(\phi; a_\phi, b_\phi). \quad (2.10)$$

Specifically,

$$\pi_U(\lambda, \rho, \phi) \propto \pi(\lambda; 1, 1) \pi(\rho; 1, 1) \pi(\phi; 1, 1), \quad (2.11)$$

$$\pi_J(\lambda, \rho, \phi) \propto \pi(\lambda; 0, 1) \pi(\rho; .5, .5) \pi(\phi; 1, .5), \quad (2.12)$$

$$\pi_R(\lambda, \rho, \phi) \propto \pi(\lambda; 0, 1) \pi(\rho; .5, .5) \pi(\phi; .5, .5). \quad (2.13)$$

The following theorems are provided to characterize a subclass of priors from (2.10), for which the associated posteriors are proper, and then find out whether the three candidate priors belong to this subclass.

Lemma 2.1 Suppose that the likelihood function and prior density are given in (2.4) and (2.10), respectively. Then the posterior distribution under this prior is proper if and only if $a_\lambda + n > 0$, $a_\rho + n_1 > 0$, $b_\rho + n_2 > 0$, $a_\phi + n_1 + n_2 > 0$, and $b_\phi + n_{12} > 0$.

Proof From (2.4) and (2.10) the joint posterior is given by

$$\pi(\lambda, \rho, \phi | \mathbf{d}) \propto \lambda^{n+a_\lambda-1} e^{-\lambda t_i} \rho^{n_1+a_\rho-1} (1-\rho)^{n_2+b_\rho-1} \phi^{n_1+n_2+a_\phi-1} (1-\phi)^{n_{12}+b_\phi-1}. \quad (2.14)$$

The propriety follows the properties of beta and gamma integrals.

Theorem 2.1 Suppose that $\{(t_i, \delta_i, \psi_i) : 1 \leq i \leq n\}$ are independent pairs from the Marshall-Olkin BVE model. Then the posterior distribution under each of non-informative priors, π_U , π_J , and π_R , is proper if and only if $n_1 \geq 1$, $n_2 \geq 1$, and $n_{12} \geq 1$.

The proof of this Theorem follows from Lemma 2.1 and (2.11)-(2.13).

From (2.14), one can recognize that λ , ρ and ϕ have independent posterior distributions of standard forms. Table 2.2 summarizes these posterior distributions under the selected noninformative priors.

Table 2.2. Posterior Distributions for the Marshall-Olkin BVE

θ	λ	ρ	ϕ
$\pi_U(\theta \mathbf{d})$	$G(n+1, t_i)$	$B(n_1+1, n_2+1)$	$B(n_1+n_2+1, n_{12}+1)$
$\pi_J(\theta \mathbf{d})$	$G(n, t_i)$	$B(n_1+\frac{1}{2}, n_2+\frac{1}{2})$	$B(n_1+n_2+1, n_{12}+\frac{1}{2})$
$\pi_R(\theta \mathbf{d})$	$G(n, t_i)$	$B(n_1+\frac{1}{2}, n_2+\frac{1}{2})$	$B(n_1+n_2+\frac{1}{2}, n_{12}+\frac{1}{2})$

2.5 Bayesian Analysis for the ACBVE Models

As pointed out earlier, the Marshall-Olkin BVE is singular since the chance of simultaneous failure is positive. This distribution has been modified in many ways

to yield nonsingular BVE's. Examples of these are Freund's ACBVE, Block-Basu's (B-B) ACBVE, Sarkar's ACBVE, and Gumbel-B's ACBVE (cf. Chapter 1). For all these models, after suitable reparameterization, the new likelihood can be presented in the form (2.1). Let (λ, ρ, ϕ) be a reparameterization for the parameter vector in the original model such that λ ($\lambda > 0$) and ρ ($0 < \rho < 1$) represent the hazard rate for T and the probability of observing $\Delta = 1$, respectively. The new parameter vector ϕ , denoting the complement of (λ, ρ) under the reparameterization, falls in the parameter space $R(\lambda, \rho)$. The relationship between θ and (λ, ρ, ϕ) , and ϕ and $R(\lambda, \rho)$ can be viewed from Table 2.3 given below.

Table 2.3. The ACBVE Reparameterization

Model	B-B & Sarkar	Gumbel-B	Freund
λ	$\lambda_1 + \lambda_2 + \lambda_{12}$	$(\lambda_1^{r-1} + \lambda_2^{r-1})^r$	$\alpha + \beta$
ρ	$\frac{\lambda_1}{\lambda_1 + \lambda_2}$	$\frac{\lambda_1^{r-1}}{\lambda_1^{r-1} + \lambda_2^{r-1}}$	$\frac{\alpha}{\alpha + \beta}$
ϕ	$\lambda_1 + \lambda_2$	r	$(\alpha' - \alpha, \beta' - \beta)$
$R(\lambda, \rho)$	$(0, \lambda)$	$(0, 1)$	$(0, \lambda - \lambda\rho) \times (0, \lambda\rho)$

The likelihood function of a ACBVE is given by the general form

$$L(\lambda, \rho, \phi|\mathbf{d}) \propto \lambda^n e^{-\lambda t} \rho^{n_1} (1 - \rho)^{n_2} \mathbf{1}_{[\phi \in R(\lambda, \rho)]}. \quad (2.15)$$

Relating (2.15) to (2.1), one sees that $\theta_1 = (\lambda, \rho)$, $\theta_{[-1]} = \phi$, $L_1(\theta_1|\mathbf{d}) = L_1(\lambda, \rho|\mathbf{d}) \propto \lambda^n e^{-\lambda t} \rho^{n_1} (1 - \rho)^{n_2}$ and $L_0(\theta_{[-1]}|\theta_1, \mathbf{d}) = L_0(\theta_{[-1]}|\mathbf{d}) \mathbf{1}_{[\theta_{[-1]} \in R(\theta_1)]} \propto \mathbf{1}_{[\phi \in R(\lambda, \rho)]}$.

This inevitable nonregularity, attributed to the nature of competing risks data and ACBVE model assumption, blocks one from finding the noninformative priors arising out of the Fisher information matrix. As a remedy, the two-stage prior elicitation strategy proposed in Section 2.2 is used to find noninformative priors. At the first stage, one assumes $\pi_{U_1} \equiv \pi_{U_1}(\phi|\lambda, \rho) \propto 1$ for $\phi \in R(\lambda, \rho)$. Then the integrated

likelihood is given by

$$\int L(\lambda, \rho, \phi | \mathbf{d}) \pi_{U_1} d\phi = L_1(\lambda, \rho | \mathbf{d}) \propto \lambda^n e^{-\lambda t} \rho^{n_1} (1 - \rho)^{n_2}. \quad (2.16)$$

With (2.16), the second-stage marginal priors are derived accordingly.

2.5.1 Marginal Jeffreys' Prior and Marginal Reference Prior

We first recognize the decomposition of $L_1(\lambda, \rho | \mathbf{d})$ as

$$L_1(\lambda, \rho | \mathbf{d}) = L_{1,c}(\lambda | \mathbf{d}) L_{1,d}(\rho | \mathbf{d}), \quad (2.17)$$

where

$$L_{1,c}(\lambda | \mathbf{d}) \propto \lambda^n e^{-\lambda t}, \quad (2.18)$$

$$L_{1,d}(\rho | \mathbf{d}) \propto \rho^{n_1} (1 - \rho)^{n_2}. \quad (2.19)$$

The Fisher information matrix, $\mathbf{I}_1 = \mathbf{I}_1(\lambda, \rho)$, derived from (2.16) has the form

$$\mathbf{I}_1(\lambda, \rho) = \begin{bmatrix} \mathbf{I}_{1,c}(\lambda) & 0 \\ 0 & \mathbf{I}_{1,d}(\rho) \end{bmatrix},$$

where $\mathbf{I}_{1,c}(\lambda) = \lambda^{-2}$ and $\mathbf{I}_{1,d}(\rho) = \rho^{-1}(1 - \rho)^{-1}$ are derived from (2.18) and (2.19), respectively. Thus, the second-stage Jeffreys' prior is given by

$$\pi_{J_2}(\lambda, \rho) = \pi_{J_2,c}(\lambda) \pi_{J_2,d}(\rho), \quad (2.20)$$

where

$$\pi_{J_2,c}(\lambda) \propto \lambda^{-1}, \quad (2.21)$$

$$\pi_{J_2,d}(\rho) \propto \rho^{-\frac{1}{2}} (1 - \rho)^{-\frac{1}{2}}. \quad (2.22)$$

Since $\mathbf{I}_1(\lambda, \rho)$ is diagonal, the second-stage reference prior is also the same as Jeffreys' prior irrespective to the choice of the parameter of interest. Thus, we propose the two-stage prior

$$\pi_{JU} \equiv \pi_{JU}(\lambda, \rho, \phi) \propto \lambda^{-1} \rho^{-\frac{1}{2}} (1 - \rho)^{-\frac{1}{2}} \mathbf{1}_{[\phi \in R(\lambda, \rho)]}. \quad (2.23)$$

Due to the factoring of the likelihood in (2.17) and the independence of λ and ρ , prior π_{JU} is thus probability matching for λ .

A second two-stage prior using Laplace's prior for both λ and ρ at the second stage is given by

$$\pi_{UU} \propto \mathbf{1}_{[\phi \in R(\lambda, \rho)]}. \quad (2.24)$$

2.5.2 Propriety of Posteriors and Bayesian Inference

Both priors π_{JU} and π_{UU} found in the previous section are improper. We must examine the propriety of the resulting posteriors. Observe the fact that $R(\lambda, \rho)$ is compact regardless of which ACBVE model is chosen. Note that both π_{JU} and π_{UU} belong to the general class of priors

$$\pi(\lambda; a_\lambda, b_\lambda) \pi(\rho; a_\rho, b_\rho) \mathbf{1}_{[\phi \in R(\lambda, \rho)]}, \quad (2.25)$$

where $\pi(\lambda; a_\lambda, b_\lambda)$ and $\pi(\rho; a_\rho, b_\rho)$ are defined in the previous section. The resulting joint posterior is given by

$$\pi(\lambda, \rho, \phi | \mathbf{d}) \propto \lambda^{n+a_\lambda-1} e^{-\lambda t_i} \rho^{n_1+a_\rho-1} (1-\rho)^{n_2+b_\rho-1} \mathbf{1}_{[\phi \in R(\lambda, \rho)]}. \quad (2.26)$$

For π_{JU} , these parameters are

$$(a_\lambda, b_\lambda; a_\rho, b_\rho) = (0, 1; .5, .5). \quad (2.27)$$

For π_{UU} , these parameters are

$$(a_\lambda, b_\lambda; a_\rho, b_\rho) = (1, 1; 1, 1). \quad (2.28)$$

The propriety of the posterior given in (2.26) is a consequence of the following theorems.

Lemma 2.2 Suppose that $a_\lambda + n > 0$, $a_\rho + n_1 > 0$, and $b_\rho + n_2 > 0$. Then the posterior given in (2.26) is proper.

Proof Integrating out (2.26) with respect to ϕ , one obtains the marginal posterior given by

$$\pi(\lambda, \rho | \mathbf{d}) \propto \lambda^{n+a_\lambda-1} e^{-\lambda t_t} \rho^{n_1+a_\rho-1} (1-\rho)^{n_2+b_\rho-1}. \quad (2.29)$$

Under the assumptions, the propriety follows from finiteness of beta and gamma integrals.

The following theorem is now a consequence of the above lemma.

Theorem 2.2 Let $\{(t_i, \delta_i) : 1 \leq i \leq n\}$ have the joint likelihood function given in (2.15). π_{JU} and π_{UU} are the two stagewise priors. The resulting posterior densities are proper if and only if $n_1 \geq 1$ and $n_2 \geq 1$.

In Table 2.4, we summarize the posterior distributions of λ and ρ . Posterior distributions for ϕ can not always be explicitly obtained. These are mixtures of standard distributions if available. Nevertheless, all posterior moments are finite and can be obtained in closed forms. These results are presented in Table 2.5.

Table 2.4. Posterior Distributions for λ and ρ of ACBVE's

θ	λ	ρ
$\pi_{UU}(\theta \mathbf{d})$	$G(n+1, t_t)$	$B(n_1+1, n_2+1)$
$\pi_{JU}(\theta \mathbf{d})$	$G(n, t_t)$	$B(n_1+\frac{1}{2}, n_2+\frac{1}{2})$

Table 2.5. Posterior Distributions (Moments) for ϕ of ACBVE's

ACBVE	B-B & Sarkar	Gumbel-B	Freund	
ϕ	$\lambda_1 + \lambda_2$	r	$d_1 = \alpha' - \alpha$	$d_2 = \beta' - \beta$
$\pi(\phi \mathbf{d})$	π_{UU}	$G^{(n)}(t_t)$	$U_{0,1}$	—
	π_{JU}	$G^{(n-1)}(t_t)$	$U_{0,1}$	—
$E(\phi \mathbf{d}, \pi)$	π_{UU}	$(n+1)/2t_t$	1/2	$q_U^{(E)}(n, n_2, t_t)$
	π_{JU}	$n/2t_t$	1/2	$q_J^{(E)}(n, n_2, t_t)$
$V(\phi \mathbf{d}, \pi)$	π_{UU}	$(n+1)(n+5)/12t_t$	1/12	$q_U^{(V)}(n, n_2, t_t)$
	π_{JU}	$n(n+4)/12t_t^2$	1/12	$q_J^{(V)}(n, n_2, t_t)$

where

$$\begin{aligned}
U_{0,1} &\sim \text{Uniform}(0, 1) \\
G^{(m)}(x) &\sim m^{-1} \sum_{i=1}^m G(i, x), \\
q_U^{(E)}(m, k, x) &= \frac{m}{m+2} \frac{k+1}{2x}, \\
q_J^{(E)}(m, k, x) &= \frac{m-1}{m+1} \frac{k+\frac{1}{2}}{2x}, \\
q_U^{(V)}(m, k, x) &= \frac{m+1}{m+2} \frac{k+1}{2x^2} \left(\frac{(m+2)(k+2)}{3(m+3)} - \frac{(m+1)(k+1)}{4(m+2)} \right), \\
q_J^{(V)}(m, k, x) &= \frac{m}{m+1} \frac{k+\frac{1}{2}}{2x^2} \left(\frac{(m+1)(k+\frac{3}{2})}{3(m+2)} - \frac{m(k+\frac{1}{2})}{4(m+1)} \right).
\end{aligned}$$

2.6 Prior Performance

As a means to evaluate the performance of the proposed noninformative priors, we propose comparing the coverage probabilities of Bayesian credible intervals with the corresponding frequentist confidence intervals for parameters λ , ρ , and ϕ (of the Marshall-Olkin BVE). Basically, we shall be looking at the closeness between the Bayesian posterior tail probability and the corresponding frequentist coverage probability for the parameter of interest.

Write $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_{[-1]})$ as before, where θ_1 is a scale parameter. Without loss of generality, let θ_1 be the parameter of interest. Define the posterior distribution function of θ_1 under prior π as

$$F_\pi(y) = P(\theta_1 < y | \mathbf{d}_n^*, \pi). \quad (2.30)$$

Also, let $\theta_{1,\pi}^{(\alpha)}(\mathbf{d}_n^*)$ denote the corresponding the posterior α -quantile of θ_1 under prior π and data $\mathbf{d}_n^*$ of size n , namely,

$$F_\pi(\theta_{1,\pi}^{(\alpha)}(\mathbf{d}_n^*)) = \alpha. \quad (2.31)$$

We say that π is a first order probability matching prior if

$$P(\theta_1 < \theta_{1,\pi}^{(\alpha)}(\mathbf{d}_n^*)|\boldsymbol{\theta}) = \alpha + o(n^{-\frac{1}{2}}), \quad (2.32)$$

while π is a second order probability matching prior if

$$P(\theta_1 < \theta_{1,\pi}^{(\alpha)}(\mathbf{d}_n^*)|\boldsymbol{\theta}) = \alpha + o(n^{-1}). \quad (2.33)$$

First, we consider the parameter λ from the BVE model. Recall that $T_t|\lambda \sim G(n, \lambda)$. From Tables 2.2 and 2.4, for a typical prior $\pi(\lambda; a_\lambda, b_\lambda)$, the posterior is $G(n + a_\lambda, t_t)$. In particular, (a_λ, b_λ) 's associated with π_U , π_J , π_R , π_{UU} , and π_{JU} equal to $(1, 1)$, $(0, 1)$, $(0, 1)$, $(1, 1)$, and $(0, 1)$, respectively. Let $\lambda_\pi^{(\alpha)}(t_t)$ denote the posterior α -quantile of λ given t_t and prior π . Thus,

$$\begin{aligned} \alpha &= P(\lambda < \lambda_\pi^{(\alpha)}(t_t)|t_t) \\ &= P(\lambda t_t < \xi_\lambda^{(\alpha)}), \end{aligned} \quad (2.34)$$

where $\xi_\lambda^{(\alpha)}$ denotes the α -quantile of the distribution $G(n + a_\lambda, 1)$. The corresponding frequentist coverage probability is given by

$$\int P(\lambda < \lambda_\pi^{(\alpha)}(T_t)|\lambda) dF_{T_t} = P(G^* < \xi_\lambda^{(\alpha)}), \quad (2.35)$$

where F_{T_t} denotes the probability distribution of T_t and $G^* \sim G(n, 1)$. The above equality suggests that the frequentist coverage probability can be affected by both α and a_λ besides the sample size n . The exception occurs only when $a_\lambda = 0$. Both (2.34) and (2.35) lead to the same value α . Then the probability matching for λ is perfect irrespective to the choice of n and α . For the BVE examples, one finds that the perfect matching for λ is achieved when π_J or π_R is selected for the Marshall-Olkin BVE, or π_{JU} is selected for the ACBVE's. Utilizing (2.35), the frequentist probabilities at selected posterior quantiles are computed in SPLUS. Theoretical and simulation results of this computation, presented in Tables 2.6 and 2.7, respectively, can then

be used as a reference for evaluating the probability matching performance. Column 3-11 provide the frequentist distributions evaluated at the posterior α -quantiles ($\alpha = .05, .2, .3, .4, .5, .6, .7, .8, .95$) given π ($\pi = \pi_U, \pi_{UU}, \pi_J, \pi_{JU}$) and samples of size n ($n = 10, 20, 50, 100$).

Table 2.6. Exact Frequentist Coverage Probabilities for λ of BVE's under π_U, π_{UU}, π_J , and π_{JU}

n	π	$\alpha = .05$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .5$	$\alpha = .6$	$\alpha = .7$	$\alpha = .8$	$\alpha = .95$
5	$\pi_U(\pi_{UU})$	.1244	.3523	.4711	.5753	.6684	.7521	.8275	.8949	.9791
	$\pi_J(\pi_{JU})$	.0500	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9500
20	$\pi_U(\pi_{UU})$	.0795	.2701	.3829	.4883	.5879	.6820	.7710	.8548	.9682
	$\pi_J(\pi_{JU})$	.0500	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9500
50	$\pi_U(\pi_{UU})$	.0670	.2427	.3514	.4556	.5561	.6530	.7466	.8364	.9625
	$\pi_J(\pi_{JU})$	.0500	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9500
100	$\pi_U(\pi_{UU})$	.0615	.2295	.3359	.4392	.5398	.6379	.7335	.8264	.9593
	$\pi_J(\pi_{JU})$	.0500	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9500

Table 2.7. Simulated Frequentist Coverage Probabilities for λ of BVE's under π_U, π_{UU}, π_J , and π_{JU}

n	π	$\alpha = .05$	$\alpha = .2$	$\alpha = .3$	$\alpha = .4$	$\alpha = .5$	$\alpha = .6$	$\alpha = .7$	$\alpha = .8$	$\alpha = .95$
5	$\pi_U(\pi_{UU})$	.1313	.3607	.4733	.5780	.6530	.7500	.8307	.8937	.9820
	$\pi_J(\pi_{JU})$	.0567	.2017	.2983	.4043	.4983	.6103	.7160	.8047	.9497
20	$\pi_U(\pi_{UU})$	.0810	.2580	.3883	.4747	.5787	.6967	.7667	.8520	.9710
	$\pi_J(\pi_{JU})$	.0507	.1897	.3047	.3870	.4970	.6087	.6893	.7933	.9513
50	$\pi_U(\pi_{UU})$	.0663	.2337	.3430	.4463	.5510	.6623	.7420	.8413	.9580
	$\pi_J(\pi_{JU})$	.0460	.1960	.2940	.3847	.4953	.6000	.6930	.8033	.9500
100	$\pi_U(\pi_{UU})$	.0627	.2320	.3363	.4480	.5507	.6253	.7333	.8223	.9610
	$\pi_J(\pi_{JU})$	.0497	.2027	.3000	.4037	.5063	.5930	.7013	.7940	.9517

As cited earlier, the matching criterion of λ is unaffected by the prior selection for ρ (or ϕ) due to the factored posterior. Except for the Laplace prior, all other selected noninformative priors perform extremely well in matching the posterior probabilities of λ with the corresponding frequentist coverage probabilities.

When the original random variables have continuous distributions, Jeffreys' prior, in the absence of nuisance parameters, is known to be a second order probability matching prior. There is no general theory supporting this feature of Jeffreys' prior

for discrete distributions. However, it still seems worthwhile to compare the Bayesian distribution for those parameters of interest (e.g., ρ and ϕ) at various points with the corresponding frequentist probabilities for moderate sample size, $n = n_1 + n_2 + n_{12}$.

In the following discussion, we shall establish the general formula for computing the frequentist probabilities, where the underlying random variables are discrete, and marginal posterior densities of model parameters are standard or posterior quantiles are accessible. Later on, BVE examples are considered for illustration.

We consider the general setup, where $\mathbf{V} \equiv (V_1, \dots, V_l)'$ follows some discrete distribution with joint pdf given by

$$f_{\mathbf{v}}(\mathbf{v}|\boldsymbol{\theta}) \equiv P(\mathbf{V} = \mathbf{v}|\boldsymbol{\theta}), \quad (2.36)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$, $\mathbf{v} = (v_1, \dots, v_l)'$, and l may differ from l' .

Let $\pi(\boldsymbol{\theta}|\mathbf{v})$ be the proper posterior density given $\mathbf{v}$ and prior $\pi(\boldsymbol{\theta})$. Suppose that the sequence of marginal posterior quantiles $\{\theta_{i,\pi}^{(\alpha)}(\mathbf{v}) : 1 \leq i \leq l'\}$ are obtainable for $0 < \alpha < 1$. Then at the first stage, for any fixed θ_i and α , the frequentist distribution evaluated at the posterior α -quantile is given by

$$\sum_{\{\mathbf{v} : \theta_i < \theta_{i,\pi}^{(\alpha)}(\mathbf{v})\}} f_{\mathbf{v}}(\mathbf{v}|\boldsymbol{\theta}). \quad (2.37)$$

In particular, let $\mathbf{V} = (N_1, N_2)$ with pdf

$$f(n_1, n_2|\rho, \phi) = \frac{n!}{n_1!n_2!n_{12}!} \rho^{n_1} (1 - \rho)^{n_2} \phi^{(n_1+n_2)} (1 - \phi)^{n_{12}}. \quad (2.38)$$

We say $(N_1, N_2) \sim \text{Mult}(n; \phi\rho, \phi(1 - \rho))$. The beta marginal posteriors of ρ and ϕ , corresponding to π_U , π_J , or π_R , have already been derived in Table 2.2. Given $\alpha \in (0, 1)$ and (n_1, n_2, n_{12}) , let $\xi_{\rho}^{(\alpha)}(n_1, n_2)$ and $\xi_{\phi}^{(\alpha)}(n_1, n_2)$ denote the α -quantile for $B(n_1 + a_{\rho}, n_2 + b_{\rho})$ and $B(n_1 + n_2 + a_{\phi}, n_{12} + b_{\phi})$, respectively. According to (2.38),

the frequentist coverage probabilities for ρ and ϕ are given by

$$\begin{aligned} & \sum_{\{(i,j): \rho < \xi_\rho^{(\alpha)}(i,j)\}} f(i, j | \rho, \phi) \\ = & \sum_{\{(i,j): \rho < \xi_\rho^{(\alpha)}(i,j)\}} \left(\frac{n!}{i!j!(n-i-j)!} \rho^i (1-\rho)^j \phi^{i+j} (1-\phi)^{(n-i-j)} \right), \end{aligned} \quad (2.39)$$

$$\begin{aligned} & \sum_{\{(i,j): \phi < \xi_\phi^{(\alpha)}(i,j)\}} f(i, j | \rho, \phi) \\ = & \sum_{\{(i,j): \phi < \xi_\phi^{(\alpha)}(i,j)\}} \left(\frac{n!}{(i+j)!(n-i-j)!} \phi^{i+j} (1-\phi)^{n-i-j} \right). \end{aligned} \quad (2.40)$$

For ACBVE's, one simply identifies the binomial distribution, $N_1 \sim \text{Bin}(n; \rho)$. From Table 2.5, $\rho | \mathbf{d} \sim B(n_1 + a_\rho, n_2 + b_\rho)$. Applying (2.37), one obtains the frequentist coverage probability

$$\sum_{\{i: \rho < \xi_\rho^{(\alpha)}(i, n-i)\}} \left(\frac{n!}{i!(n-i)!} \rho^i (1-\rho)^{n-i} \right). \quad (2.41)$$

These numerical results will be presented with the transformed Q-Q plots by using the SPLUS software package. We convert the Q-Q plot axes from quantiles into the corresponding probabilities. Figures 2.1-2.3 and 2.4-2.6 demonstrate the exact frequentist-versus-posterior transformed Q-Q plots for ρ and ϕ , respectively, based on $(N_1, N_2) \sim \text{Mult}(n; \phi\rho, \phi(1-\rho))$ with the combination of $n = 10, 50, 100$, and $(\rho, \phi) = (.75, .2), (.5, .4), (.33, .6), (.125, .8)$.

It follows from Figures 2.1-2.6 that the frequentist probabilities of ρ or ϕ given in (2.39)-(2.41) under selected noninformative priors do depend on the values of parameters and choices of α .

Due to the discreteness of the underlying distribution, the matching concept cannot be justified rigorously. Our general finding is that when Jeffreys' rule is applied,

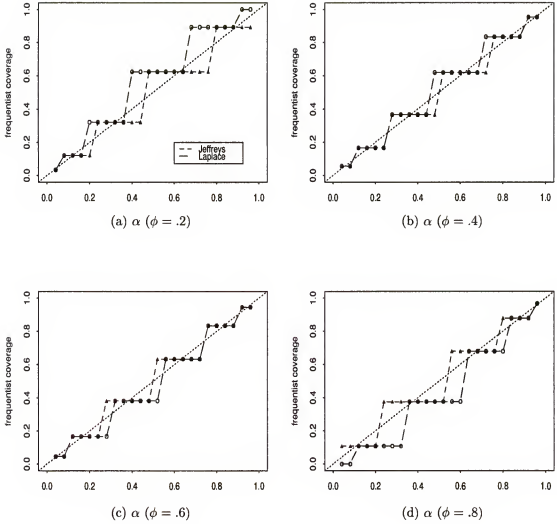


Figure 2.1. Transformed Q-Q Plots: Posterior Probabilities of ϕ vs Frequentist Probabilities of ϕ under Jeffreys' and Laplace's priors and samples of size 10. (a) transformed Q-Q plot for $\phi = .2$; (b) transformed Q-Q plot for $\phi = .4$; (c) transformed Q-Q plot for $\phi = .6$; (d) transformed Q-Q plot for $\phi = .8$.

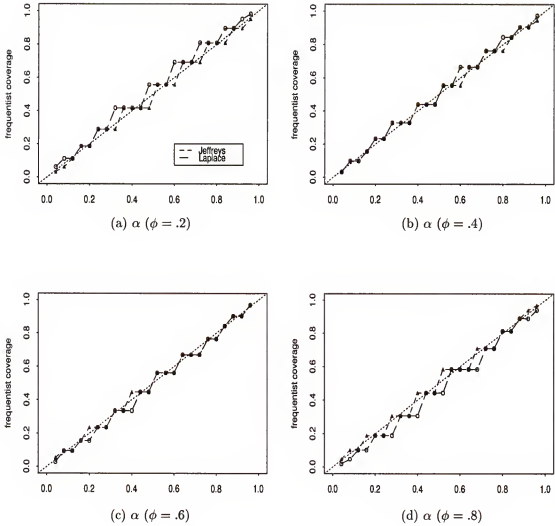


Figure 2.2. Transformed Q-Q Plots: Posterior Probabilities of ϕ vs Frequentist Probabilities of ϕ under Jeffreys' and Laplace's priors and samples of size 50. (a) transformed Q-Q plot for $\phi = .2$; (b) transformed Q-Q plot for $\phi = .4$; (c) transformed Q-Q plot for $\phi = .6$; (d) transformed Q-Q plot for $\phi = .8$.

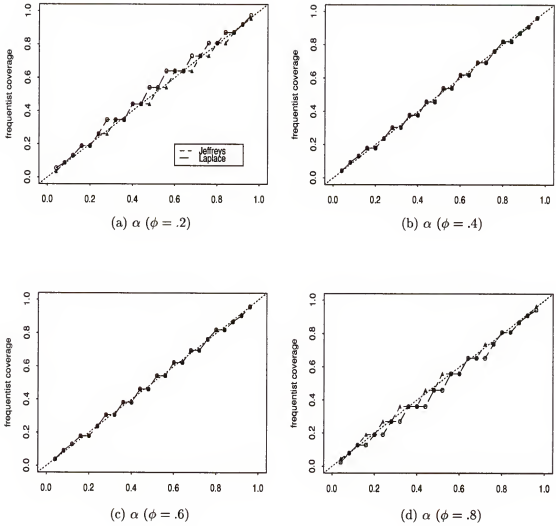


Figure 2.3. Transformed Q-Q Plots: Posterior Probabilities of ϕ vs Frequentist Probabilities of ϕ under Jeffreys' and Laplace's priors and samples of size 100. (a) transformed Q-Q plot for $\phi = .2$; (b) transformed Q-Q plot for $\phi = .4$; (c) transformed Q-Q plot for $\phi = .6$; (d) transformed Q-Q plot for $\phi = .8$.

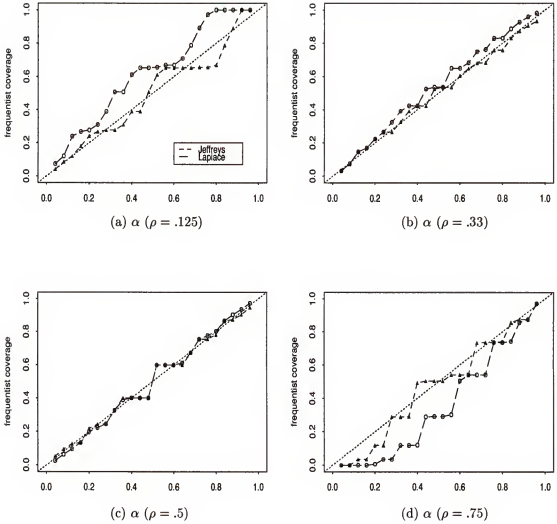


Figure 2.4. Transformed Q-Q Plots: Posterior Probabilities of ρ vs Frequentist Probabilities of ρ under Jeffreys' and Laplace's priors, and samples of size 10. (a) transformed Q-Q plot for $\rho = .125$; (b) transformed Q-Q plot for $\rho = .33$; (c) transformed Q-Q plot for $\rho = .5$; (d) transformed Q-Q plot for $\rho = .75$.

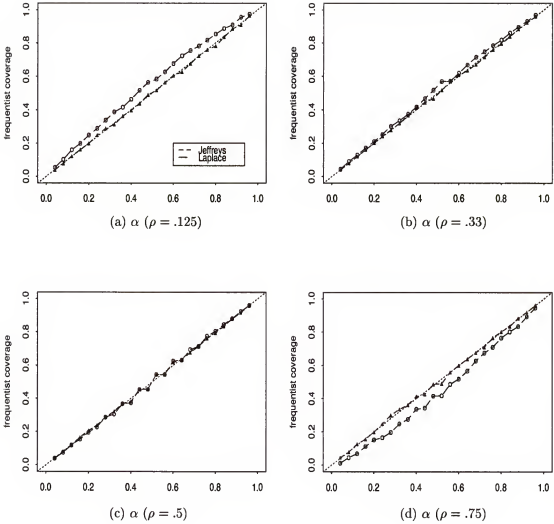


Figure 2.5. Transformed Q-Q Plots: Posterior Probabilities of ρ vs Frequentist Probabilities of ρ under Jeffreys' and Laplace's priors, and samples of size 50. (a) transformed Q-Q plot for $\rho = .125$; (b) transformed Q-Q plot for $\rho = .33$; (c) transformed Q-Q plot for $\rho = .5$; (d) transformed Q-Q plot for $\rho = .75$.

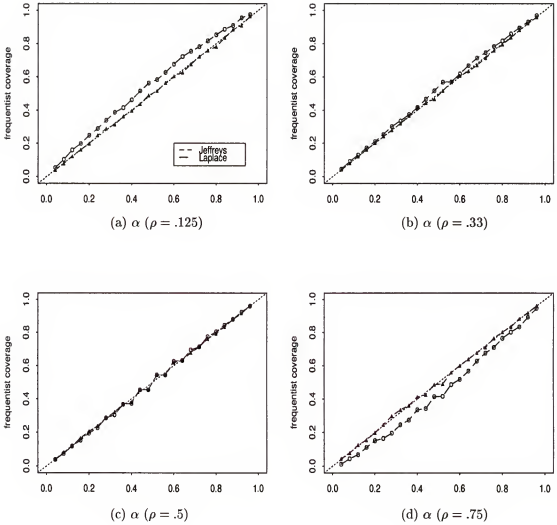


Figure 2.6. Transformed Q-Q Plots: Posterior Probabilities of ρ vs Frequentist Probabilities of ρ under Jeffreys' and Laplace's priors, and samples of size 100. (a) transformed Q-Q plot for $\rho = .125$; (b) transformed Q-Q plot for $\rho = .33$; (c) transformed Q-Q plot for $\rho = .5$; (d) transformed Q-Q plot for $\rho = .75$.

either on a regular basis, or in the two-stage specification, one usually obtains a relatively better matching property in comparison with Laplace's rule. If one formally extends the matching concept from the continuous case to the discrete case, then π_J and π_{JU} are indeed second order matching priors.

Tail posterior percentiles, associated with $\alpha = .05$ and $\alpha = .95$, are fairly close to their frequentist counterparts, which are robust to parameter values and the non-informative priors that are selected. This suggests that one could rely on the 90% Bayesian credible set to achieve the corresponding frequentist coverage probabilities for ρ and ϕ .

It appears that π_J and π_U (for the Marshall-Olkin BVE), or π_{JU} and π_{UU} (for ACBVE's) have more similar converge probabilities for parameters, ρ and ϕ , in the central part ($\rho = .33, .5, \phi = .4, .6$) than those near the boundary ($\rho = .125, .75, \phi = .2, .8$).

2.7 Identifiable ACBVE via Informative Priors

The Bayesian posterior distribution for ϕ (or ϕ) varies among different ACBVE assumptions. For instance, the posterior distribution of ϕ is exactly the same as the prior when the Gumbel-B ACBVE is assumed. For other ACBVE's, it is a weighted result of prior and the likelihood function. The Bayesian estimator of ϕ is uniquely determined and essentially is the midpoint of $R(\lambda, \rho)$ when the noninformative prior is assumed. The prior information of ϕ may occasionally be reflected from history (e.g., medical data, social science census). The strength of Bayesian methodology in resolving the nonidentifiability allows one to adopt the prior information to provide reasonable inference. To see this, we conduct a simulation study for the Block-Basu

ACBVE, where ϕ follows a conditional beta prior which is extended from the first-stage uniform prior $B(1, 1)$

$$\pi(\phi|\lambda) \propto \left(\frac{\phi}{\lambda}\right)^{a-1} \left(1 - \frac{\phi}{\lambda}\right)^{b-1}. \quad (2.42)$$

Without loss of generality, we let $\lambda = 1$. Priors $B(4, 16)$, $B(20, 80)$, $B(10, 10)$, $B(50, 50)$, $B(14, 6)$, and $B(70, 30)$ are assumed. Since the posterior of ϕ here is independent of ρ and δ , the samples of sizes 10 and 50 are generated from $T \sim G(1, 1)$. The posterior means and standard deviations of ϕ are summarized in Table 2.8.

Table 2.8. Posterior Estimates for ϕ under three-parameter ACBVE's and Beta Priors

$\pi(\phi \lambda)$	$E_{\pi}(\phi)$	$E_{\pi}(\phi \mathbf{d})(V_{\pi}^{1/2}(\phi \mathbf{d}))$	
		$n = 10$	$n = 50$
$B(4, 16)$	.2	.2183 (.1213)	.2043 (.0945)
$B(20, 80)$	.2	.2278 (.0862)	.2042 (.0502)
$B(10, 10)$	.5	.5550 (.2098)	.5088 (.1330)
$B(50, 50)$	.5	.5543 (.1782)	.5086 (.0882)
$B(14, 6)$	.7	.8011 (.2480)	.7142 (.1413)
$B(70, 30)$	.7	.8146 (.2281)	.7066 (.1076)

Figure 2.7 provides a more detailed insight of the simulation results. In each of the sub-plots, a specified prior distribution of ϕ is drawn along with the associated posterior distributions of ϕ given samples of sizes 10 and 50. These graphs suggest that priors eventually dominated posteriors as sample size increases. Bayesians guide the inference towards the truth even when the likelihood is not fully identified.

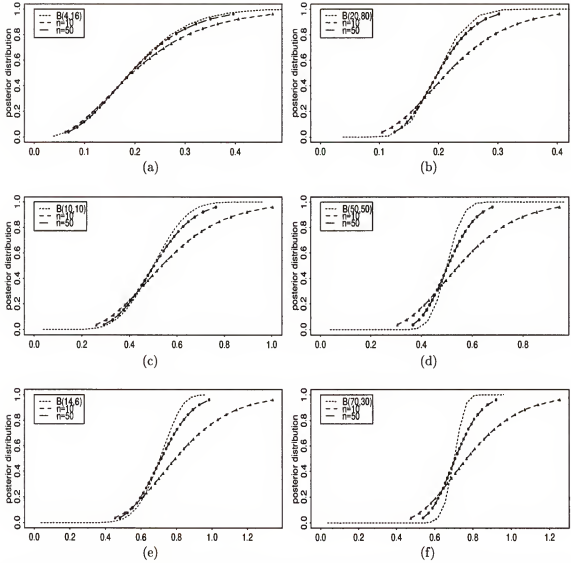


Figure 2.7. Beta Prior Distributions and the Associated Posterior Distributions of ϕ under $n = 10$ and $n = 50$: (a) prior: $B(4, 16)$; (b) prior: $B(20, 80)$; (c) prior: $B(10, 10)$; (d) prior: $B(50, 50)$; (e) prior: $B(14, 6)$; (f) prior: $B(70, 30)$.

CHAPTER 3

BAYESIAN ANALYSIS OF GENERALIZED BIVARIATE EXPONENTIAL MODELS

3.1 Introduction

In lifetime analyses, there are circumstances where auxiliary information (covariates), such as external forces or characteristics of individuals, may have impacts on lifetime. In this chapter, a class of generalized BVE models is introduced, in which (T, Δ, Ψ) is adjusted with covariate information. To distinguish, we shall, from now on, refer to the BVE models (excluding covariates) as baseline BVE models. Essentially, the generalized linear model technique is used for the construction of BVE models with covariates. This technique is applied to the exponential variate, T , and binary variates, Δ and Ψ . Along with the choice of the canonical link in GLM, the generalized BVE models are obtained as combinations of log-linear models and logit-linear models. In the terminology of survival analysis, these are Cox's proportional hazard models and proportional odds models. Unlike the baseline BVE models, feasible inclusion of covariates allows for dependence between T and (Δ, Ψ) by adding in common factors shared between the two. Random censoring is assumed. Denoting the censoring variable by C , all we observe is $\min(T, C)$ besides (Δ, Ψ) . Bayesian methodology with noninformative priors is employed for statistical analyses based on the marginal likelihood function associated with (T, Δ, Ψ) . Laplace's flat prior and Jeffreys' prior, each suitably modified by the stagewise specification, are selected as candidates. The organization of this chapter is as follows. The generalized BVE models are constructed in Section 3.2. In Sections 3.3 and 3.4, prior densities and the associated posterior densities are derived for ACBVE's and the Marshall-Olkin BVE,

respectively. The propriety of posteriors is examined under the criteria of Ibrham and Laud (1991) and Shao and Chen (1998). Section 3.5 discusses the strength of modified noninformative priors in the context of factorial experiment design. In Section 3.6, the Gibbs sampling and Metropolis-Hasting sampling methods are used for the generation of marginal posteriors. We illustrate our findings with a lung cancer clinical trial dataset in Section 3.7.

3.2 Generalized BVE Models

In the general setting, each observation (T, Δ, Ψ) is now concatenated with a covariate vector and a censoring indicator $\mathbf{1}_c$ assumed to be independent of (T, Δ, Ψ) . Let $\mathbf{z}_i$ ($k \times 1$) denote the covariate vector for the i^{th} observation, and write $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_{n_+}]'$ for the $n_+ \times k$ design matrix and $\mathbf{Z}_n = [\mathbf{z}_1, \dots, \mathbf{z}_n]'$ which will be assumed of rank k throughout. Instead of an explicit use of a censoring variable, we distinguish censored observations from failures via the subscripts. Specifically, indices, $1, \dots, n$, will mark the failures, while indices, $n+1, \dots, n_+$, will mark censored individuals. The data for i^{th} individual are now expressed by the quadruplets $\mathbf{d}_i \equiv (t_i, \delta_i, \psi_i; \mathbf{z}_i)$. Let $\mathbf{d} \equiv \{\mathbf{d}_i : 1 \leq i \leq n_+\}$. The contribution to the likelihood from the censoring variable C does not depend on the parameters associated with (T, Δ, Ψ) , which is given by $\prod_{i=n+1}^{n_+} P(T_i > t_i)$. Thus, the marginal likelihood function of (T, Δ, Ψ) is validly used for statistical inference, including in the derivation of noninformative priors.

In this section, we restrict ourselves to the Marshall-Olkin BVE. The generalized linear model (GLM; cf. McCullagh and Nelder, 1996) technique relates the mean of outcome variables of the regular exponential family to covariates through the “link function”. This technique is appropriately extended to the model generalization of T , Δ , and Ψ . One common choice of the link function is the canonical link. The model

structures are given by

$$\log(\lambda(\mathbf{z})) = \mathbf{z}'\boldsymbol{\beta}, \quad (3.1)$$

$$\text{logit}(\rho(\mathbf{z})) = \mathbf{z}'\boldsymbol{\gamma}, \quad (3.2)$$

$$\text{logit}(\phi(\mathbf{z})) = \mathbf{z}'\boldsymbol{\alpha}, \quad (3.3)$$

where $\mathbf{z} = (1, z_2, \dots, z_k)'$. Baseline BVE models can be viewed as GLM's when the mean of the response does not depend on covariates, i.e., $\lambda = e^{\beta_1}$, $\rho = e^{\gamma_1}/1 + e^{\gamma_1}$, and $\phi = e^{\alpha_1}/1 + e^{\alpha_1}$. Physical interpretations for regression coefficients β_j 's, γ_j 's, and α_j 's may vary from one text to another.

In view of the nature of the parameters λ , ρ , and ϕ in the BVE, one may adjust these parameters with covariates via hierarchical modeling. Let $\lambda(\mathbf{z})$, $\rho(\mathbf{z})$, and $\phi(\mathbf{z})$ denote the covariate-adjusted functions for λ , ρ , and ϕ , respectively. As given in (3.1)-(3.3), both Cox's proportional hazard model and the logistic regression model are adopted in the competing risks context for modeling $\lambda(\mathbf{z})$ and $\rho(\mathbf{z})$ (or $\phi(\mathbf{z})$), respectively.

3.3 Generalization of ACBVE Models

Based on the GLM technique with the canonical link, the class of generalized ACBVE models (cf. Section 2.4) are generated by assuming (3.1) and (3.2). Parameters $\boldsymbol{\phi}$ appear in the same way as in the baseline model. Then the marginal likelihood function for realization $\{(t_i, \delta_i; \mathbf{z}_i) : 1 \leq i \leq n_+\}$ is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi} | \mathbf{d}) \propto \prod_{i=1}^n \left\{ e^{\mathbf{z}_i' \boldsymbol{\beta}} \frac{(e^{\mathbf{z}_i' \boldsymbol{\gamma}})^{\delta_i}}{1 + e^{\mathbf{z}_i' \boldsymbol{\gamma}}} \right\} \prod_{i=1}^{n_+} \left(\exp(-e^{\mathbf{z}_i' \boldsymbol{\beta}} t_i) \right) \mathbf{1}_{[\boldsymbol{\phi} \in R(\boldsymbol{\beta}, \boldsymbol{\gamma})]}. \quad (3.4)$$

3.3.1 Noninformative Priors and Posteriors

To find noninformative priors on the basis of (3.4), we employ the stagewise elicitation as the Fisher information matrix can be not fully specified. At the first stage, we propose Laplace's prior, i.e., $\pi_{U_1} \equiv \pi_{U_1}(\phi|\beta, \gamma) \propto 1$ for $\phi \in R(\beta, \gamma)$, some specified compact set. Thus, the integrated likelihood function $L_1(\beta, \gamma|\mathbf{d})$ is given by

$$\begin{aligned} & L_1(\beta, \gamma|\mathbf{d}) \\ &= \int L(\beta, \gamma, \phi|\mathbf{d}) \pi_{U_1}(\phi|\beta, \gamma) d\phi \\ &\propto \prod_{i=1}^n \left\{ e^{\mathbf{z}'_i \beta} \frac{(e^{\mathbf{z}'_i \gamma})^{\delta_i}}{(1 + e^{\mathbf{z}'_i \gamma})} \right\} \prod_{i=1}^{n+} \left(\exp(-e^{\mathbf{z}'_i \beta} t_i) \right). \end{aligned} \quad (3.5)$$

At the second stage, Laplace's prior, $\pi_{U_2} \equiv \pi_{U_2}(\beta, \gamma) \propto 1$, and Jeffreys' prior, $\pi_{J_2} \equiv \pi_{J_2}(\beta, \gamma)$, are computed from (3.5). Suppose that $\pi_{UU} \equiv \pi_{UU}(\beta, \gamma, \phi) \propto \pi_{U_1}(\phi|\beta, \gamma) \pi_{U_2}(\beta, \gamma)$ is specified. The corresponding joint posterior density, denoted by $\pi_{UU}(\beta, \gamma, \phi|\mathbf{d})$, can be factored as $\pi_{U_1}(\phi|\beta, \gamma) \pi_{UV}(\beta|\mathbf{d}) \pi_{UV}(\gamma|\mathbf{d})$, where

$$\pi_{UV}(\beta|\mathbf{d}) \propto \left(\prod_{i=1}^n e^{\mathbf{z}'_i \beta} \right) \prod_{i=1}^{n+} \left(\exp(-e^{\mathbf{z}'_i \beta} t_i) \right), \quad (3.6)$$

and

$$\pi_{UV}(\gamma|\mathbf{d}) \propto \prod_{i=1}^n \left(\frac{(e^{\mathbf{z}'_i \gamma})^{\delta_i}}{1 + e^{\mathbf{z}'_i \gamma}} \right). \quad (3.7)$$

Next, the Fisher information matrix derived from (3.5) is given by

$$\mathbf{I}_1(\beta, \gamma) = \begin{bmatrix} n_+^{-1} \mathbf{Z}' \mathbf{Z} & \mathbf{0}_{k \times k} \\ \mathbf{0}_{k \times k} & n^{-1} \mathbf{Z}'_n \mathbf{M}(\gamma) \mathbf{Z}_n \end{bmatrix},$$

where $\mathbf{M}(\gamma)$ is a $n \times n$ diagonal matrix with the i^{th} diagonal element given by $M_i(\gamma) = e^{\mathbf{z}'_i \gamma} / (1 + e^{\mathbf{z}'_i \gamma})^2$. Thus, Jeffreys' prior is obtained as $\pi_{J_2} \equiv \pi_{J_2}(\beta, \gamma) \propto$

$[\|\mathbf{Z}'\mathbf{Z}\|\|\mathbf{Z}'_n\mathbf{M}(\boldsymbol{\gamma})\mathbf{Z}_n\|]^{\frac{1}{2}}$. And the posterior under $\pi_{JU} \equiv \pi_{JU}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}) \propto \pi_{U_1}(\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{\gamma})\pi_{J_2}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is given by

$$\pi_{JU}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}|\mathbf{d}) \propto \pi_{U_1}(\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{\gamma})\pi_{UU}(\boldsymbol{\beta}|\mathbf{d})\pi_{JU}(\boldsymbol{\gamma}|\mathbf{d}), \quad (3.8)$$

where $\pi_{UU}(\boldsymbol{\beta}|\mathbf{d})$ is given in (3.6), and

$$\pi_{JU}(\boldsymbol{\gamma}|\mathbf{d}) \propto \prod_{i=1}^n \left(\frac{(e^{\mathbf{z}'_i \boldsymbol{\gamma}})^{\delta_i}}{1 + e^{\mathbf{z}'_i \boldsymbol{\gamma}}} \right) |\mathbf{Z}'_n \mathbf{M}(\boldsymbol{\gamma}) \mathbf{Z}_n|^{\frac{1}{2}}. \quad (3.9)$$

To compute the determinant, $|\mathbf{Z}'_n \mathbf{M}(\boldsymbol{\gamma}) \mathbf{Z}_n|$, one may refer to the following result attributed to Cauchy and Binet (cf. Noble and James, 1977).

Cauchy-Binet Result Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]'$ be a $n \times k$ ($k \leq n$) matrix of rank k and $\mathbf{M}^*$ be a $n \times n$ diagonal matrix with the i^{th} diagonal element M_i^* . Denote $c(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}) = \|\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\|^2$, for $1 \leq i_1 < \dots < i_k \leq n$, and $P_n^{(k)} = \{(i_1, \dots, i_k) : 1 \leq i_1 < \dots < i_k \leq n, c(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}) > 0\}$. Then

$$|\mathbf{A}'\mathbf{M}^*\mathbf{A}| = \sum_{P_n^{(k)}} c(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}) \left(\prod_{u=1}^k M_{i_u}^* \right). \quad (3.10)$$

By (3.10),

$$|\mathbf{Z}'_n \mathbf{M}(\boldsymbol{\gamma}) \mathbf{Z}_n| = \left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \left(\prod_{u=1}^k M_{i_u}(\boldsymbol{\gamma}) \right) \right\}. \quad (3.11)$$

Thus,

$$\pi_{JU}(\boldsymbol{\gamma}|\mathbf{d}) \propto \prod_{i=1}^n \left(\frac{(e^{\mathbf{z}'_i \boldsymbol{\gamma}})^{\delta_i}}{1 + e^{\mathbf{z}'_i \boldsymbol{\gamma}}} \right) \left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \left(\prod_{u=1}^k M_{i_u}(\boldsymbol{\gamma}) \right) \right\}^{\frac{1}{2}}. \quad (3.12)$$

3.3.2 Propriety of Posteriors

The immediate task is to ascertain the propriety of the posterior densities (3.4) and (3.8). Since $\pi_{U_1}(\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{\gamma})$ is proper for any given $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, we may proceed directly

to the marginal posterior densities

$$\pi_{UU}(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{d}) = \pi_{UU}(\boldsymbol{\beta} | \mathbf{d}) \pi_{UV}(\boldsymbol{\gamma} | \mathbf{d}), \quad (3.13)$$

and

$$\pi_{JU}(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{d}) = \pi_{UU}(\boldsymbol{\beta} | \mathbf{d}) \pi_{JU}(\boldsymbol{\gamma} | \mathbf{d}). \quad (3.14)$$

The convergence of $\int \pi_{UU}(\boldsymbol{\beta} | \mathbf{d}) d\boldsymbol{\beta}$ follows from a theorem of Ibrahim and Laud (1991), which provides sufficient conditions for the propriety of Jeffreys' prior. For completeness, the theorem is stated below.

Theorem (Ibrahim and Laud, 1991) Suppose $Y_1, \dots, Y_n$ are independent variates with pdf's

$$f(y_i | \theta_i, \omega_i, \phi) = \exp\{\omega_i / \eta(y_i \theta_i - b(\theta_i)) + g(y_i, \eta)\} \quad (3.15)$$

where $\theta_i = \theta(\mathbf{z}'_i \boldsymbol{\beta})$ and η and $\omega_1, \dots, \omega_n$ are known constants. If the pdf's are bounded functions of the θ_i 's and $\mathbf{Z}_n = [\mathbf{z}_1, \dots, \mathbf{z}_n]'$ is of full column rank, then under Jeffreys' prior, a sufficient condition for the posterior distribution of $\boldsymbol{\beta}$ to be proper is that

$$\int_{S_\theta} \exp\{(\omega/\eta)(y\theta - b(\theta))\} \left(\frac{d^2 b(\theta)}{d\theta^2} \right)^{1/2} d\theta < \infty, \quad (3.16)$$

where S_θ denotes for the integral domain of θ .

To prove the finiteness of $\int \pi_{UU}(\boldsymbol{\beta} | \mathbf{d}) d\boldsymbol{\beta}$ analytically, we begin with the inequality

$$\pi_{UU}(\boldsymbol{\beta} | \mathbf{d}) \leq \prod_{i=1}^n \left(e^{\mathbf{z}'_i \boldsymbol{\beta}} \exp(-e^{\mathbf{z}'_i \boldsymbol{\beta}} t_i) \right). \quad (3.17)$$

One may note that the right-hand side of (3.17) corresponds to (3.15) via

- $f(t_i) = \exp(-e^{\mathbf{z}'_i \boldsymbol{\beta}} t_i + \mathbf{z}'_i \boldsymbol{\beta})$,
- $\eta = \omega_1 = \dots = \omega_n = 1$,
- $\theta_i = -\exp(\mathbf{z}'_i \boldsymbol{\beta})$, and

- $b(\theta_i) = -\log(-\theta_i)$ and $S_\theta = (-\infty, 0)$.

Since Jeffreys' prior, proportional to $|\mathbf{Z}'_n \mathbf{Z}_n|^{\frac{1}{2}}$, is a constant, one can verify (3.16) from

$$\int_{-\infty}^0 \exp(t\theta + \log(-\theta)) (1/\theta^2)^{\frac{1}{2}} d\theta = \int_0^{\infty} e^{-t\theta} d\theta = t^{-1} < \infty.$$

The next theorem examines the propriety of $\pi_{JU}(\boldsymbol{\gamma})$.

Theorem 3.1 Let $\{\Delta_i : 1 \leq i \leq n\}$ be independent binary outcomes with $P(\Delta_i = 1|\mathbf{z}_i) = e^{\mathbf{z}'_i \boldsymbol{\gamma}} / (1 + e^{\mathbf{z}'_i \boldsymbol{\gamma}}) = 1 - P(\Delta_i = 0|\mathbf{z}_i)$. Suppose that Jeffreys' prior is assumed, which is given by $\pi_J(\boldsymbol{\gamma}) \propto |\mathbf{Z}'_n \mathbf{M}(\boldsymbol{\gamma}) \mathbf{Z}_n|^{\frac{1}{2}}$, where $\mathbf{M}(\boldsymbol{\gamma})$ is defined in Section 3.3.1. Then $\pi_J(\boldsymbol{\gamma})$ is a proper density if the matrix $\mathbf{Z}_n$ is of full column rank. Therefore, the associated posterior density, equivalent to (3.12), is a proper posterior, i.e.,

$$\int \pi_{JU}(\boldsymbol{\gamma}|\mathbf{d}) d\boldsymbol{\gamma} < \infty.$$

Proof We shall show that Jeffreys' prior in this case is proper, which implies the propriety of the corresponding posterior.

Apply Jensen's inequality to $\pi_J(\boldsymbol{\gamma})$. Then

$$\left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \prod_{u=1}^k \left(\frac{e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}}}{(1 + e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}})^2} \right) \right\}^{\frac{1}{2}} \leq \sum_{P_n^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})]^{\frac{1}{2}} \left\{ \prod_{u=1}^k \left(\frac{(e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}})^{\frac{1}{2}}}{1 + e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}}} \right) \right\}.$$

Next, we examine the finiteness of integrals of the type

$$[c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})]^{\frac{1}{2}} \int \prod_{u=1}^k \left(\frac{(e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}})^{\frac{1}{2}}}{1 + e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}}} \right) d\boldsymbol{\gamma}. \quad (3.18)$$

By change-of-variables $r_u = \mathbf{z}'_{i_u} \boldsymbol{\gamma}$ and $x_u = e^{r_u/2}$, then

$$[c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})]^{\frac{1}{2}} \int \prod_{u=1}^k \left(\frac{(e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}})^{\frac{1}{2}}}{1 + e^{\mathbf{z}'_{i_u} \boldsymbol{\gamma}}} \right) d\boldsymbol{\gamma} \quad (3.19)$$

$$= \prod_{u=1}^k \left(\int_{-\infty}^{\infty} \frac{(e^{r_u})^{\frac{1}{2}}}{1 + e^{r_u}} dr_u \right) \quad (3.20)$$

$$= \prod_{u=1}^k \left(\int_0^{\infty} \frac{dx_u}{(1 + x_u^2)} \right) \quad (3.21)$$

$$< \infty.$$

The result follows.

We shall then appeal to the result of Shao and Chen (1998) for verifying the propriety of $\pi_{UU}(\boldsymbol{\gamma}|\mathbf{d})$.

Theorem (Shao and Chen, 1998) Suppose that $Y_1, \dots, Y_n$ are independent binary response variables with pdf's $P(Y_i = 1) = F(\mathbf{z}'_i \boldsymbol{\gamma}) = 1 - P(Y_i = 0)$, where $F(\cdot)$ is some probability distribution function. Define $\mathbf{U}^* = \text{Diag}[u_1^*, \dots, u_n^*]$ and $\mathbf{Z}_n^* = \mathbf{Z}'_n \mathbf{U}^*$, where y_i is the realization of Y_i and $u_i^* = (-1)^{y_i}$. Then given Laplace's prior, the joint posterior of $\boldsymbol{\gamma}$ is given by $\prod_{i=1}^n \{[F(\mathbf{z}'_i \boldsymbol{\gamma})]^{y_i} [1 - F(\mathbf{z}'_i \boldsymbol{\gamma})]^{1-y_i}\}$. This posterior density is proper if and only if:

(C1) $\mathbf{Z}_n$ is of full rank,

(C2) there exists a *positive vector* $\mathbf{v} \in R^n$, i.e., each component $v_i > 0$, such that $[\mathbf{Z}^*]'\mathbf{v} = \mathbf{0}$, and

(C3) $\int_{-\infty}^{\infty} |y|^k dF(y) < \infty$.

Furthermore, (C3) holds for logit, probit, complementary log and certain t-links.

Thus, if in addition to (C1), (C2) is assumed with $u_i^* = (-1)^{\delta_i}$ and $\mathbf{Z}_n^* = \mathbf{Z}'_n \mathbf{U}^*$, then convergence of $\int \pi_{UU}(\boldsymbol{\gamma}|\mathbf{d}) d\boldsymbol{\gamma}$ follows since F is the logistic distribution in the present case. In summary, to ensure the propriety of posteriors, one needs to assume (C1) when the prior is π_{JU} , and both (C1) and (C2) when the prior is π_{UU} .

3.4 Generalization of the Marshall-Olkin BVE Model

In order to include covariate information in the Marshall-Olkin BVE model, we adopt the modeling schemes, (3.1)-(3.3), and the marginal likelihood approach described in Section 3.2. Given $\{(t_i, \delta_i, \psi_i; \mathbf{z}_i) : 1 \leq i \leq n_+\}$, the marginal likelihood function is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) \propto \prod_{i=1}^n \left\{ e^{\mathbf{z}_i' \boldsymbol{\beta}} \left(\frac{(e^{\mathbf{z}_i' \boldsymbol{\gamma}})^{\delta_i}}{1 + e^{\mathbf{z}_i' \boldsymbol{\gamma}}} \right)^{\psi_i} \left(\frac{(e^{\mathbf{z}_i' \boldsymbol{\alpha}})^{\psi_i}}{1 + e^{\mathbf{z}_i' \boldsymbol{\alpha}}} \right) \right\} \prod_{i=1}^{n_+} \left(\exp(-e^{\mathbf{z}_i' \boldsymbol{\beta}} t_i) \right). \quad (3.22)$$

3.4.1 Noninformative Priors

Typical noninformative priors are computed based on (3.22). Consistent with Section 3.3, Laplace's prior and Jeffreys' prior are regarded as candidates. Under Laplace's prior, the joint posterior density is simply proportional to (3.22), which is given by the expression

$$\pi_U(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) = \pi_U(\boldsymbol{\beta} | \mathbf{d}) \pi_U(\boldsymbol{\gamma} | \mathbf{d}) \pi_U(\boldsymbol{\alpha} | \mathbf{d}), \quad (3.23)$$

where

$$\pi_U(\boldsymbol{\beta} | \mathbf{d}) \propto \left(\prod_{i=1}^n e^{\mathbf{z}_i' \boldsymbol{\beta}} \right) \prod_{i=1}^{n_+} \left(\exp(-e^{\mathbf{z}_i' \boldsymbol{\beta}} t_i) \right), \quad (3.24)$$

$$\pi_U(\boldsymbol{\gamma} | \mathbf{d}) \propto \prod_{i=1}^n \left\{ \frac{(e^{\mathbf{z}_i' \boldsymbol{\gamma}})^{\delta_i}}{1 + e^{\mathbf{z}_i' \boldsymbol{\gamma}}} \right\}^{\psi_i}, \quad (3.25)$$

$$\pi_U(\boldsymbol{\alpha} | \mathbf{d}) \propto \prod_{i=1}^n \left\{ \frac{(e^{\mathbf{z}_i' \boldsymbol{\alpha}})^{\psi_i}}{1 + e^{\mathbf{z}_i' \boldsymbol{\alpha}}} \right\}. \quad (3.26)$$

Next, let

$$\mathbf{M}_1(\mathbf{v}) = \text{Diag}[M_{1,1}(\mathbf{v}), \dots, M_{1,n}(\mathbf{v})], \quad (3.27)$$

and

$$\mathbf{M}_2(\mathbf{v}) = \text{Diag}[M_{2,1}(\mathbf{v}), \dots, M_{2,n}(\mathbf{v})], \quad (3.28)$$

where

$$M_{1,i}(\mathbf{v}) = \frac{e^{\mathbf{z}'_i \mathbf{v}}}{(1 + e^{\mathbf{z}'_i \mathbf{v}})^2}, \quad (3.29)$$

$$M_{2,i}(\mathbf{v}) = \frac{e^{\mathbf{z}'_i \mathbf{v}}}{1 + e^{\mathbf{z}'_i \mathbf{v}}}. \quad (3.30)$$

The Fisher information matrix is given by

$$\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) = \begin{bmatrix} n_+^{-1} \mathbf{Z}' \mathbf{Z} & \mathbf{0}_{k \times k} & \mathbf{0}_{k \times k} \\ \mathbf{0}_{k \times k} & n^{-1} \mathbf{Z}'_n [\mathbf{M}_1(\boldsymbol{\gamma}) \mathbf{M}_2(\boldsymbol{\alpha})] \mathbf{Z}_n & \mathbf{0}_{k \times k} \\ \mathbf{0}_{k \times k} & \mathbf{0}_{k \times k} & n^{-1} \mathbf{Z}'_n \mathbf{M}_1(\boldsymbol{\alpha}) \mathbf{Z}_n \end{bmatrix}.$$

Then Jeffreys' prior is given by

$$\pi_J \equiv \pi_J(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \propto |\mathbf{Z}'_n [\mathbf{M}_1(\boldsymbol{\gamma}) \mathbf{M}_2(\boldsymbol{\alpha})] \mathbf{Z}_n|^{1/2} |\mathbf{Z}'_n \mathbf{M}_1(\boldsymbol{\alpha}) \mathbf{Z}_n|^{1/2}. \quad (3.31)$$

By the determinant expansion given in (3.10), one can rewrite (3.31) as

$$\begin{aligned} \pi_J(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) &\propto \left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \left(\prod_{u=1}^k M_{1,i_u}(\boldsymbol{\gamma}) M_{2,i_u}(\boldsymbol{\alpha}) \right) \right\}^{\frac{1}{2}} \\ &\times \left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \left(\prod_{u=1}^k M_{1,i_u}(\boldsymbol{\alpha}) \right) \right\}^{\frac{1}{2}}. \end{aligned} \quad (3.32)$$

Accordingly, the posterior density is given by

$$\pi_J(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) = \pi_U(\boldsymbol{\beta} | \mathbf{d}) \pi_J(\boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}), \quad (3.33)$$

where $\pi_U(\boldsymbol{\beta} | \mathbf{d})$ is the same as in (3.24), and

$$\pi_J(\boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) \propto \prod_{i=1}^n \left\{ \left(\frac{(e^{\mathbf{z}'_i \boldsymbol{\gamma}})^{\delta_i}}{1 + e^{\mathbf{z}'_i \boldsymbol{\gamma}}} \right)^{\psi_i} \left(\frac{(e^{\mathbf{z}'_i \boldsymbol{\alpha}})^{\psi_i}}{1 + e^{\mathbf{z}'_i \boldsymbol{\alpha}}} \right) \right\}$$

$$\begin{aligned}
& \times \left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \left(\prod_{u=1}^k M_{1,i_u}(\gamma) M_{2,i_u}(\alpha) \right) \right\}^{\frac{1}{2}} \\
& \times \left\{ \sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \left(\prod_{u=1}^k M_{1,i_u}(\alpha) \right) \right\}^{\frac{1}{2}}. \tag{3.34}
\end{aligned}$$

3.4.2 Propriety of Posteriors

In order to check the propriety of posteriors, one needs to verify the convergence of $\int \pi_U(\gamma|\mathbf{d}) d\gamma$, $\int \pi_U(\alpha|\mathbf{d}) d\alpha$, and $\int \pi_J(\gamma, \alpha|\mathbf{d}) d\gamma d\alpha$.

Shao and Chen's (1998) result is applied to the evaluation of the first two integrals. Let $\mathbf{Z}_\Psi$, $(n_1 + n_2) \times k$, be comprised of row vectors $\{\mathbf{z}'_i : 1 \leq i \leq n \text{ and } \psi_i = 1\}$. Then $\int \pi_U(\gamma|\mathbf{d}) d\gamma$ converges if $\mathbf{Z}_\Psi$ is of rank k and $\mathbf{Z}_\Psi^* \mathbf{v}_\Psi = \mathbf{0}$ for some positive vector $\mathbf{v}_\Psi$, $(n_1 + n_2) \times 1$, where $\mathbf{Z}_\Psi^* = [\mathbf{Z}_\Psi]'\mathbf{U}^*$ and $u_i^* = (-1)^{\delta_i}$. The convergence of $\int \pi_U(\alpha|\mathbf{d}) d\alpha$ holds if $\mathbf{Z}_n$ is of rank k and $[\mathbf{Z}_n']\mathbf{v}_n = \mathbf{0}$ for some positive vector $\mathbf{v}_n$, $n \times 1$, where $\mathbf{Z}_n^* = \mathbf{Z}_n'\mathbf{U}^*$ and $u_i^* = (-1)^{\psi_i}$.

To justify the propriety of (3.34), we consider the inequalities

$$\begin{aligned}
c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) & \leq c(\mathbf{z}_1^*, \dots, \mathbf{z}_k^*) = \max\{c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})\}, \\
M_{2,i}(\alpha) & \leq 1,
\end{aligned}$$

for all $(i_1, \dots, i_k) \in P_n^{(k)}$ and $1 \leq i \leq n$. It follows that

$$\begin{aligned}
& \int \pi_J(\gamma, \alpha|\mathbf{d}) d\gamma d\alpha \\
& \leq c(\mathbf{z}_1^*, \dots, \mathbf{z}_k^*) \left\{ \sum_{P_n^{(k)}} \int \prod_{u=1}^k \left(\frac{(e^{\mathbf{z}'_{i_u} \gamma})^{\frac{1}{2}}}{1 + e^{\mathbf{z}'_{i_u} \gamma}} \right) d\gamma \right\} \\
& \times \left\{ \sum_{P_n^{(k)}} \int \prod_{u=1}^k \left(\frac{(e^{\mathbf{z}'_{i_u} \alpha})^{\frac{1}{2}}}{1 + e^{\mathbf{z}'_{i_u} \alpha}} \right) d\alpha \right\}. \tag{3.35}
\end{aligned}$$

Repeating the steps (3.19)-(3.21) in the proof of Theorem 3.1, the convergence is attained if $\mathbf{Z}_n$ satisfies rank k assumption.

Similar conclusions can be drawn as in the ACBVE cases. The propriety is established under Laplace's prior if both (C1) and (C2) hold in the way stated above. The calculation based on Jeffreys' prior requires (C1) only.

3.5 Application to a Sample with Categorical Covariates

For many statistical applications, covariates (or explanatory variables) are categorical taking on a finite number of values. In these applications, the task of finding posteriors in Sections 3.3 and 3.4 is greatly simplified. Primarily, we consider a factorial experiment design in the current context. Suppose that certain interpretable parameters such as the main effects or interaction effects among the factors are of interest. In the following discussion, a variety of probability matching properties are investigated among linear functions of β_j 's when Laplace's or Jeffreys' rule is applied to a complete dataset.

3.5.1 Defining the Design Matrix

In a factorial experiment design, the underlying sample can be classified into k mutually exclusive categories (all possible combinations of the factor levels). The covariate vector for individual i , $\mathbf{z}_i = (z_{i,1}, \dots, z_{i,k})'$ is defined such that $z_{i,j}$ is the indicator variable for category j , for $1 \leq j \leq k$. Thus, the design matrix $\mathbf{Z}$ is a composition of the unit vectors.

3.5.2 Likelihood Functions and Posteriors

Let $n_{1j} = \sum_{i=1}^n \delta_i \psi_i \mathbf{1}_{[z_{i,j}=1]}$, $n_{2j} = \sum_{i=1}^n (1 - \delta_i) \psi_i \mathbf{1}_{[z_{i,j}=1]}$, $n_{12j} = \sum_{i=1}^n (1 - \psi_i) \mathbf{1}_{[z_{i,j}=1]}$, $n_{\cdot j} = n_{1j} + n_{2j} + n_{12j}$, and $t_{tj} = \sum_{i=1}^n t_i \mathbf{1}_{[z_{i,j}=1]}$. Then one can simplify the likelihood functions (3.4) of ACBVE, and (3.22) of the Marshall-Olkin BVE to

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi} | \mathbf{d}) \propto \prod_{j=1}^k \left\{ \exp[n_{\cdot j} \beta_j - e^{\beta_j} t_{tj}] \right\} \prod_{j=1}^k \left\{ \frac{(e^{\gamma_j})^{n_{1j}}}{(1 + e^{\gamma_j})^{n_{\cdot j}}} \right\} \mathbf{1}_{[\boldsymbol{\phi} \in R(\boldsymbol{\beta}, \boldsymbol{\gamma})]}, \quad (3.36)$$

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) &\propto \prod_{j=1}^k \left\{ \exp[n_{\cdot j} \beta_j - e^{\beta_j} t_{tj}] \right\} \\ &\times \prod_{j=1}^k \left\{ \frac{(e^{\gamma_j})^{n_{1j}}}{(1 + e^{\gamma_j})^{n_{1j} + n_{2j}}} \frac{(e^{\alpha_j})^{n_{1j} + n_{2j}}}{(1 + e^{\alpha_j})^{n_{\cdot j}}} \right\}. \end{aligned} \quad (3.37)$$

Accordingly, the priors π_{JU} , and π_J given in (3.9) and (3.32) reduce to

$$\pi_{JU}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}) \propto \prod_{j=1}^k \left(\frac{(e^{\gamma_j})^{\frac{1}{2}}}{1 + e^{\gamma_j}} \right) \mathbf{1}_{[\boldsymbol{\phi} \in R(\boldsymbol{\beta}, \boldsymbol{\gamma})]}, \quad (3.38)$$

$$\pi_J(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \propto \prod_{j=1}^k \left\{ \left(\frac{(e^{\gamma_j})^{\frac{1}{2}}}{1 + e^{\gamma_j}} \right) \left[\frac{(e^{\alpha_j})^{\frac{1}{2}}}{(1 + e^{\alpha_j})^{\frac{1}{2}}} \right] \right\}. \quad (3.39)$$

The corresponding posteriors under π_{UU} , π_{JU} , π_U , and π_J are

$$\begin{aligned} \pi_{UU}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi} | \mathbf{d}) &\propto \prod_{j=1}^k \left\{ \exp[n_{\cdot j} \beta_j - e^{\beta_j} t_{tj}] \right\} \\ &\times \prod_{j=1}^k \left\{ \frac{(e^{\gamma_j})^{n_{1j}}}{(1 + e^{\gamma_j})^{n_{\cdot j}}} \right\} \mathbf{1}_{[\boldsymbol{\phi} \in R(\boldsymbol{\beta}, \boldsymbol{\gamma})]}, \end{aligned} \quad (3.40)$$

$$\begin{aligned} \pi_{JU}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi} | \mathbf{d}) &\propto \prod_{j=1}^k \left\{ \exp[n_{\cdot j} \beta_j - e^{\beta_j} t_{tj}] \right\} \\ &\times \prod_{j=1}^k \left\{ \frac{(e^{\gamma_j})^{n_{1j} + \frac{1}{2}}}{(1 + e^{\gamma_j})^{n_{\cdot j} + 1}} \right\} \mathbf{1}_{[\boldsymbol{\phi} \in R(\boldsymbol{\beta}, \boldsymbol{\gamma})]}, \end{aligned} \quad (3.41)$$

$$\begin{aligned}
\pi_U(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) &\propto \prod_{j=1}^k \left\{ \exp[n_{\cdot j} \beta_j - e^{\beta_j} t_{t_j}] \right\} \\
&\times \prod_{j=1}^k \left\{ \frac{(e^{\gamma_j})^{n_{1j}}}{(1 + e^{\gamma_j})^{n_{1j} + n_{2j}}} \frac{(e^{\alpha_j})^{n_{1j} + n_{2j}}}{(1 + e^{\alpha_j})^{n_{\cdot j}}} \right\}, \tag{3.42}
\end{aligned}$$

$$\begin{aligned}
\pi_J(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{d}) &\propto \prod_{j=1}^k \left\{ \exp[n_{\cdot j} \beta_j - e^{\beta_j} t_{t_j}] \right\} \\
&\times \prod_{j=1}^k \left\{ \frac{(e^{\gamma_j})^{n_{1j} + \frac{1}{2}}}{(1 + e^{\gamma_j})^{n_{1j} + n_{2j} + 1}} \frac{(e^{\alpha_j})^{n_{1j} + n_{2j} + \frac{1}{2}}}{(1 + e^{\alpha_j})^{n_{\cdot j} + \frac{3}{2}}} \right\}. \tag{3.43}
\end{aligned}$$

Using the transformations: $\{\lambda_j = e^{\beta_j} : 1 \leq j \leq k\}$, $\{\rho_j = \frac{e^{\gamma_j}}{1 + e^{\gamma_j}} : 1 \leq j \leq k\}$, and $\{\phi_j = \frac{e^{\alpha_j}}{1 + e^{\alpha_j}} : 1 \leq j \leq k\}$, the marginal posterior distributions are given by

$$\lambda_j | \mathbf{d}, \pi_{JU} (\pi_{UU} \pi_J \pi_U) \sim G(n_{\cdot j}, t_{t_j}), \tag{3.44}$$

$$\rho_j | \mathbf{d}, \pi_{JU} (\pi_J) \sim B(n_{1j} + .5, n_{2j} + .5), \tag{3.45}$$

$$\rho_j | \mathbf{d}, \pi_{UU} (\pi_U) \sim B(n_{1j}, n_{2j}), \tag{3.46}$$

$$\phi_j | \mathbf{d}, \pi_{JU} (\pi_J) \sim B(n_{1j} + n_{2j} + .5, n_{12_j} + 1), \tag{3.47}$$

$$\phi_j | \mathbf{d}, \pi_{UU} (\pi_U) \sim B(n_{1j} + n_{2j}, n_{12_j}). \tag{3.48}$$

The posteriors for β_j 's, γ_j 's, and α_j 's can be obtained via one-to-one transformation.

3.5.3 Contrasts of Interest: Linear Functions of β_j 's

Consider the one-to-one reparameterization $\boldsymbol{\theta} = \mathbf{A}^{-1} \boldsymbol{\beta}$ for some nonsingular matrix $\mathbf{A}^{-1}$. The contrasts $\{\theta_1, \dots, \theta_k\}$ may particularly correspond to the ratios of hazard rates of adjacent categories in the logarithm scale. Write $\mathbf{A}' = [\mathbf{a}_1, \dots, \mathbf{a}_k]$.

Due to the factored likelihood function, posterior marginals of θ_j 's are found as follows. The marginal likelihood function of θ is given by

$$L(\theta|\mathbf{d}) \propto \exp\left\{\sum_{j=1}^k [n_j \mathbf{a}'_j \theta - \exp(\mathbf{a}'_j \theta) t_j]\right\}. \quad (3.49)$$

The corresponding Fisher information matrix is

$$I(\theta) = n_+^{-1} \mathbf{A}'[\mathbf{Z}'\mathbf{Z}]\mathbf{A}.$$

Hence, $\pi_J(\theta) \propto 1$, and $\mathbf{I}^{-1}(\theta) = (\mathbf{I}^{ij}(\theta))$ is a constant matrix independent of θ . Using the formulae (1.36), (1.38), and (1.40), one can conclude that both the second order optimal and HPD-based matching properties for θ_j 's hold under π_{UU} , π_{JU} , π_U , and π_J . The proofs are deferred to Appendix. In order to verify the matching properties numerically, we shall demonstrate two simulation results in Section 3.7.

3.6 Sampling Schemes

In Bayesian framework, the posterior density tightly anchors inference. As one has seen in Sections 3.3 and 3.4, one may be confronted with very complicated multivariate posterior densities so that inference requires high-dimensional numerical integration. A more attractive alternative is to conduct numerical integration via the Markov Chain Monte Carlo (MCMC) methods, in particular, the Gibbs sampling method. This requires generating samples from the conditional distributions of the parameter given the other parameters and data. Many of these conditional densities are non-standard, and one needs to employ the Metropolis-Hasting algorithm to generate samples from these conditionals.

3.6.1 Gibbs Sampling

The Gibbs sampler (cf. Gelfand and Smith (1990), Casella and George (1992), Dellaportas and Smith (1993)) is a Markovian updating scheme enabling one to obtain

samples from a joint distribution via iterated sampling from full conditional distributions. The merit of this sampling scheme is that one can obtain samples from the target marginal density avoiding tedious multi-dimensional integration. Suppose that the joint density of $(Y_1, \dots, Y_m)$ is proportional to $p(y_1, \dots, y_m)$. Let $f_1(\cdot), \dots, f_m(\cdot)$ denote the marginal density for $Y_1, \dots, Y_m$ accordingly. Suppose that one aims to sample from $f_i(\cdot)$. Ideally, it could be found by integrating out the joint density, say $f(y_1, \dots, y_m)$ with respect to all the arguments but y_i . In practice, it may not always be feasible, particularly for high-dimensional densities known only up to a multiplier constant. With Gibbs algorithm, one only needs to recognize the conditional densities (up to the normalized constant)

$$p_i(y_i | y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_m) \propto p(y_1, \dots, y_m),$$

for $1 \leq i \leq m$. Set initial vector $(y_1^{(0)}, \dots, y_m^{(0)})$. This Markovian updating procedure is stated below.

Generate $y_1^{(1)}$ from the conditional density

$$p_1(y_1 | y_2^{(0)}, \dots, y_m^{(0)}) / \int p_1(y_1 | y_2^{(0)}, \dots, y_m^{(0)}) dy_1.$$

Next, generate $y_2^{(1)}$ from the conditional pdf $p_2(y_2 | y_1^{(1)}, y_3^{(0)}, \dots, y_m^{(0)})$. Proceeding this way, we complete eventually the first iteration. In the $(j+1)^{th}$ iteration, $\mathbf{y}^{(j)} = (y_1^{(j)}, \dots, y_m^{(j)})$ is replaced for the initial vector for proceeding $\mathbf{y}^{(j+1)}$. Under mild conditions, as shown by Geman and Geman (1984) for discrete distributions, we have

$$y_i^{(j)} \sim f_i(y_i),$$

for fairly large J . However, quite often some conditionals $p_i(y_i | \mathbf{y}_{-i})$ cannot be recognized as any standard density. In order to generate samples from these non-standard pdf's, one employs the Metropolis-Hasting algorithm.

3.6.2 Metropolis-Hasting Algorithm

In connection with Gibbs sampling, the Metropolis-Hasting algorithm (cf. Chib and Greenberg, 1995) becomes useful when some of the conditional pdf's are not standard, and are known up to a multiplicative constant. For instance, $f(y^*|y) \propto p(y^*)$. To generate samples for Y from this pdf, one needs to specify a *candidate generating density*, denoted by $g(x, y)$ with $\int g(x, y)dx = 1$. Define the *probability of transition* as

$$\alpha(y, x) = \min \left\{ \frac{p(y)g(x, y)}{p(x)g(y, x)}, 1 \right\}.$$

Initialize with $y = y^{(0)}$. The algorithm is described as follows.

- Generate y^* from $g(\cdot, y^{(0)})$ and u from $U(0, 1)$.
- If $u \leq \alpha(y^{(0)}, y^*)$, then set $y^{(1)} = y^*$, else update $y^{(0)}$ by y^* .
- Repeat the previous two steps.

In this way, we generate samples $y^{(1)}, y^{(2)}, \dots$, from the marginal pdf of Y . Typically, the *candidate generating density* is a standard distribution. Among those well-developed selection tools, we shall employ the one suggested by Chib and Greenberg (1992). The scheme requires $p(y) \propto c(y)m(y)$, where $c(y)$ is uniformly bounded, and $m(y)$ is a density of standard form. Then set $g(y^*, y) = m(y^*)$ to draw candidate samples. Accordingly, $\alpha(y, y^*)$ reduces to $\min \left\{ \frac{c(y^*)}{c(y)}, 1 \right\}$.

3.7 Illustrated Examples

3.7.1 Simulation Studies

Two simulations are conducted for $k = 2$ and $k = 3$ to verify the matching properties stated in Section 3.5. The procedure is described below. First, the model parameters β and matrix $\mathbf{A}$ are selected in advance. Gibbs sampling scheme is used

to generate posterior samples for $\theta_1, \dots, \theta_k$, respectively. The simulation study is based on 5000 Markovian iterations with 1000 samples generated in each iteration. In each iteration, 5%, 25%, 50%, 75%, and 95% posterior percentiles for θ_j 's are evaluated. For each θ_j , the empirical frequentist distributions are evaluated at the five posterior percentiles. Given a specific percentage α , samples $\mathbf{d}$, and prior π , the frequentist probability $P(\theta_i < \theta_{i,\pi}^{(\alpha)}(\mathbf{d})|\theta)$ is estimated by computing the proportion of these probabilities among the 5000 iterations, which is observed to be less than α .

Example 1 For $k = 2$, we let $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (1, 0.5)'$ and select a number of (n_1, n_2) . Suppose the contrasts of interest are $\theta_1 = \beta_1 - \beta_2$, and $\theta_2 = \beta_2$ so that

$$\boldsymbol{\theta} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \boldsymbol{\beta}.$$

Here

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

Table 3.1 presents the results of the lower tail frequentist coverage probabilities of θ_i at $\theta_{i,\pi}^{(\alpha)}(\mathbf{d})$'s, where $\pi = \pi_{JU}, \pi_{UU}, \pi_J, \pi_U$, $\alpha = .05, .25, .5, .75, .95$, and $\mathbf{d}$ are the simulated data.

Table 3.1. Frequentist Tail Probabilities of θ_1 and θ_2 under $\pi_{JU}, \pi_{UU}, \pi_J$, or π_U , and $(\beta_1, \beta_2) = (1, 0.5)$

(n_1, n_2)	α	.05	.25	.50	.75	.95
(2,2)	θ_1	0.0480	0.2498	0.4948	0.7428	0.9470
	θ_2	0.0526	0.2496	0.4936	0.7536	0.9482
(2,4)	θ_1	0.0490	0.2446	0.4944	0.7524	0.9502
	θ_2	0.0540	0.2552	0.5026	0.7544	0.9524
(5,5)	θ_1	0.0514	0.2444	0.4960	0.7488	0.9462
	θ_2	0.0534	0.2586	0.5012	0.7504	0.9506
(10,10)	θ_1	0.0528	0.2422	0.4904	0.7399	0.9506
	θ_2	0.0500	0.2560	0.4986	0.7446	0.9494

Example 2 For $k = 3$, we let $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)' = (1, 0.5, 2)'$ and select a number of (n_1, n_2, n_3) . The contrasts of interest are $\theta_1 = \beta_1 - \beta_2$, $\theta_2 = \beta_2 - \beta_3$, and $\theta_3 = \beta_3$ so

that

$$\boldsymbol{\theta} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\beta}.$$

Similar to Example 1, a variety of (n_1, n_2, n_3) are selected to examine the matching property. The frequentist coverage probabilities of θ_i 's are calculated in Table 3.2.

Table 3.2. Frequentist Tail Probabilities of θ_1 , θ_2 and θ_3 under π_{JU} , π_{UU} , π_J , or π_U , and $(\beta_1, \beta_2, \beta_3) = (1, 0.5, 2)$

(n_1, n_2, n_3)	α	.05	.25	.50	.75	.95
(2,2,2)	θ_1	0.0520	0.2602	0.5060	0.7560	0.9504
	θ_2	0.0448	0.2434	0.4886	0.7370	0.9452
	θ_3	0.0532	0.2494	0.5158	0.7636	0.9542
(8,5,2)	θ_1	0.0478	0.2488	0.4938	0.7390	0.9486
	θ_2	0.0526	0.2464	0.4882	0.7638	0.9500
	θ_3	0.0492	0.2588	0.5180	0.7462	0.9518
(10,10,10)	θ_1	0.0470	0.2550	0.5080	0.7570	0.9524
	θ_2	0.0482	0.2508	0.4930	0.7608	0.9464
	θ_3	0.0510	0.2400	0.4994	0.7442	0.9514

3.7.2 Data Analysis

Next, we illustrate Bayesian procedure for generalized ACBVE's with a lung cancer clinical trial that was conducted by the Eastern Cooperative Oncology Group (ECOG). In the ECOG study, 194 patients with squamous cell carcinoma were included, of which 83 patients failed with local spread of disease (cause 1), 44 failed with metastatic spread of disease, while the remaining 67 observations were censored. Three covariates are considered: X_1 = performance status (ambulatory=0, non-ambulatory=1), X_2 = treatment ($A = 0$, $B = 1$), and X_3 = age in years. Frequentist analysis have been performed by Lagakos (1978) who assumed a cause-specific model, and by Wada, Sen and Shimakara (1995), who assumed the generalized Sarkar ACBVE. In the following discussion, Bayesian analyses under noninformative priors are provided for this study, in which covariate effects on survival are of primary

interest. The first analysis (A1) was conducted by including all three covariates given Laplace's uniform prior. In this case, Jeffreys' prior becomes computationally problematic. In the second analysis (A2), covariate X_3 (age) is excluded since its effect is found to be insignificant. In this case, both Jeffreys' prior and Laplace's prior can be used.

3.7.2.1 Analysis 1

In (A1), the covariate vector is given by $(1, X_1, X_2, X_3)'$, which is associated with coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ in the proportional hazard model, and $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)'$ in the logistic regression model, respectively. M-H algorithm within Gibbs sampling is adopted to access posterior information. To verify the convergence of the Gibbs sampler, the *potential scale reduction factor*, denoted by R , proposed by Gelmen and Rubin (Gelman and Rubin, 1992) is computed. R -values for $\beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2, \gamma_3$, and γ_4 , based on posterior samples generated from five distinct initial points, are given by 1.0068, 1.0002, 1.0003, 1.0074, 1.0064, 1.0005, 1.0020, and 1.0069, respectively. This confirms, to a certain extent, the stability of our Gibbs sampling process. This can also be revealed from the marginal posterior distributions given in Figures 3.1 and 3.2.

Denote the α posterior quantiles by q_α . In Table 3.3, we report posterior mean (B.E.), standard deviation (S.D.), and quantiles $q_{.025}, q_{.05}, q_{.25}, q_{.5}, q_{.75}, q_{.95}$, and $q_{.975}$ given the assumption of Laplace's prior.

In examining the goodness-of-fit property, we adopt the method by Gelman, Carlin, Stern, and Rubin (1995). The main idea, analogous to the standard p -value justification, is to compare the observed outcome to its posterior predictive distribution, which is given by $p_y = P(Y_{pred} > y_{obs} | \theta_{post})$, where Y_{pred} denotes the posterior predicted variable of observation y_{obs} , and is assumed to be continuous. It reflects lack-of-fit if this upper tail probability is close to 0 or 1. In practice, p_y is computed

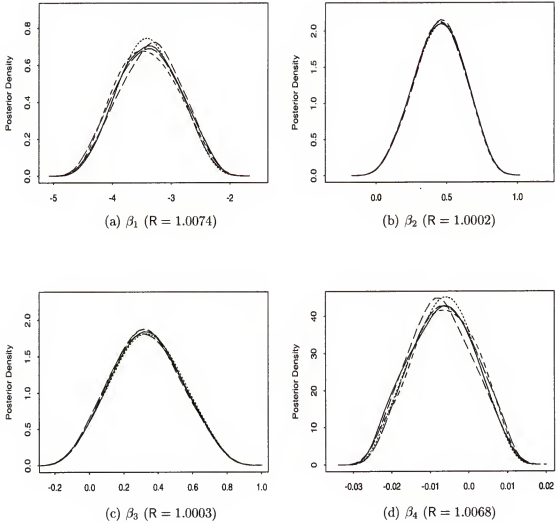


Figure 3.1. Posterior Densities of β_1 , β_2 , β_3 , and β_4 generated from five distinct initial points via Gibbs-Metropolis sampling scheme and the associated *potential scale reduction factors*, R 's: (a) posterior densities of β_1 ; (b) posterior densities of β_2 ; (c) posterior densities of β_3 ; (d) posterior densities of β_4 .

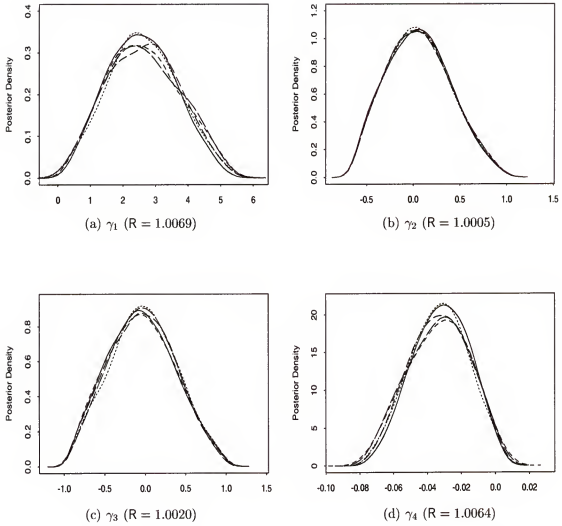


Figure 3.2. Posterior Densities of γ_1 , γ_2 , γ_3 , and γ_4 generated from five distinct initial points via Gibbs-Metropolis sampling scheme and the associated *potential scale reduction factors*, R 's: (a) posterior densities of γ_1 ; (b) posterior densities of γ_2 ; (c) posterior densities of γ_3 ; (d) posterior densities of γ_4 .

Table 3.3. ECOG Posterior Estimates under Main Effect Model and π_{UU}

θ	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	γ_4
B.E.	-3.353	.451	.328	-.0071	2.550	.057	-.035	-.032
S.D.	.513	.163	.193	.0084	1.097	.341	.422	.018
$q_{.025}$	-4.296	.133	-.040	-.0223	.593	-.533	-.798	-.061
$q_{.050}$	-4.158	.175	.005	-.021	.783	-.480	-.717	-.057
$q_{.250}$	-3.743	.336	.189	-.0134	1.722	-.192	-.349	-.0426
$q_{.500}$	-3.382	.455	.326	-.0068	2.471	.044	-.049	-.0301
$q_{.750}$	-2.980	.572	.464	-.0007	3.236	.286	.254	-.0182
$q_{.950}$	-2.523	.715	.653	.0063	4.172	.621	.669	-.0047
$q_{.975}$	-2.424	.753	.702	.0085	4.455	.730	.777	-.0016

empirically by counting the proportion of y_{new} 's generated from posterior samples θ_{post} that exceeds y_{obs} . More specifically, let $f_{\theta}(\cdot)$ denote the pdf of Y_{pred} . Suppose that the posterior samples for θ are given by $\{\theta^{(1)}, \dots, \theta^{(m)}\}$, and $y^{(1)}, \dots, y^{(m)}$ generated respectively from $f_{\theta^{(1)}}, \dots, f_{\theta^{(m)}}$ are the predicted samples. The posterior predictive distribution is given by

$$p_y = \frac{1}{m} \sum_{j=1}^m \mathbf{1}_{[y < y^{(j)}]}. \quad (3.50)$$

Using this analogy, we assess the goodness-of-fit for a binary outcome y by

$$p_y = \frac{1}{m} \sum_{j=1}^m \mathbf{1}_{[y=y^{(j)}]}. \quad (3.51)$$

This probability should be interpreted as *posterior predictive matching probability*. Thus, the higher the probability, the better the prediction (matching).

Two scatter plots for p_{t_i} 's and p_{δ_i} 's are drawn in Figure 3.3, and the corresponding empirical predictive coverage probabilities, $\sum_{i=1}^n p_{t_i}/n$ and $\sum_{i=1}^n p_{\delta_i}/n$, are computed in the right-hand corner of each plot.

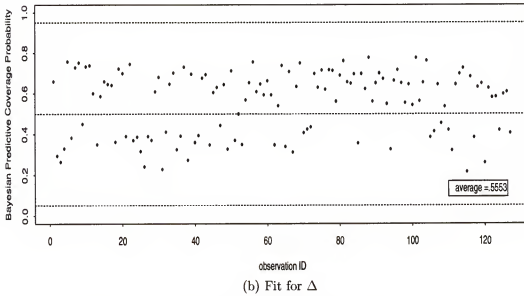
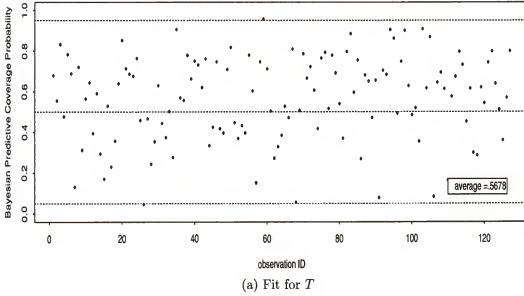


Figure 3.3. Goodness-of-Fit Plot for Failure Observations under Analysis 1: (a) Plotting Bayesian Predictive Coverage Probabilities for T_i 's; (b) Plotting Bayesian Predictive Coverage Probabilities for Δ_i 's.

3.7.2.2 Analysis 2

From a practical viewpoint, the previous analysis suggests that that “age” factor has rather insignificant effect. In the second approach, only X_2 (the ambulatory indicator) and X_3 (the treatment B indicator) are included. Thus, the two studied covariates now are categorical. Following the description in Section 3.5, ECOG samples are classified into four covariate combinations. The covariate vector for individual i is defined as $\mathbf{z}_i = (z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4})'$ such that $z_{i,1} = \mathbf{1}_{[(X_2, X_3)=(1,1)]}$, $z_{i,2} = \mathbf{1}_{[(X_2, X_3)=(1,0)]}$, $z_{i,3} = \mathbf{1}_{[(X_2, X_3)=(0,1)]}$, and $z_{i,4} = \mathbf{1}_{[(X_2, X_3)=(0,0)]}$. Posterior quantities, B.E., S.D., $q_{.025}$, $q_{.05}$, $q_{.25}$, $q_{.75}$, $q_{.95}$, and $q_{.975}$ for β_j and γ_j are presented in Table 3.4. The Bayesian model diagnosis scatter plots are given in Figure 3.4.

Table 3.4. ECOG Posterior Estimates under Interaction Model and π_{JU} , π_{UU}

θ	π	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	γ_4
B.E.	π_{JU}	-2.930	-3.577	-3.505	-3.572	0.509	1.069	0.696	0.400
S.D.	π_{JU}	0.160	0.298	0.127	0.260	0.326	0.668	0.271	0.525
B.E.	π_{UU}	-2.908	-3.493	-3.488	-3.503	0.501	0.998	0.687	0.382
S.D.	π_{UU}	0.157	0.288	0.130	0.256	0.322	0.620	0.270	0.514
$q_{.025}$	π_{JU}	-3.216	-4.181	-3.744	-4.102	-0.118	-0.117	0.167	-0.607
	π_{UU}					-0.113	-0.071	0.175	-0.613
$q_{.05}$	π_{JU}	-3.159	-4.069	-3.699	-4.004	-0.019	0.065	0.250	-0.445
	π_{UU}					-0.012	0.120	0.258	-0.446
$q_{.25}$	π_{JU}	-2.987	-3.744	-3.563	-3.717	0.289	0.639	0.507	0.050
	π_{UU}					0.300	0.726	0.518	0.066
$q_{.50}$	π_{JU}	-2.872	-3.535	-3.472	-3.532	0.506	1.061	0.689	0.396
	π_{UU}					0.520	1.176	0.702	0.424
$q_{.75}$	π_{JU}	-2.761	-3.339	-3.383	-3.355	0.727	1.513	0.875	0.750
	π_{UU}					0.744	1.663	0.889	0.792
$q_{.95}$	π_{JU}	-2.606	-3.074	-3.259	-3.117	1.055	2.238	1.150	1.284
	π_{UU}					1.077	2.458	1.168	1.349
$q_{.975}$	π_{JU}	-2.557	-2.992	-3.220	-3.043	1.164	2.499	1.242	1.466
	π_{UU}					1.188	2.748	1.466	1.539

The linear contrasts of β_j 's are considered, particularly, the relative hazard rates between categories in the logarithm scale (or linear contrasts of β_j 's). Their Bayesian posterior estimates and quantiles under π_{UU} (or π_{JU}) are presented in Table 3.5.

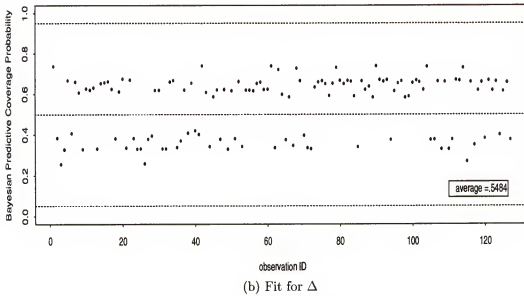
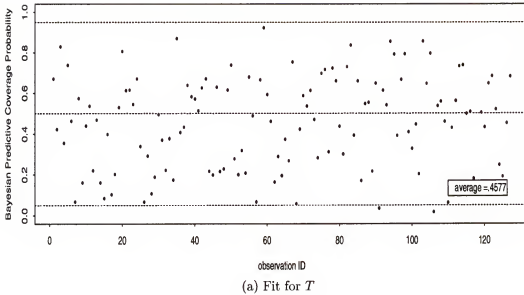


Figure 3.4. Goodness-of-Fit Plot for Failure Observations under Analysis 2: (a) Plotting Bayesian Predictive Coverage Probabilities for T_i 's; (b) Plotting Bayesian Predictive Coverage Probabilities for Δ_i 's.

Table 3.5. Posterior Estimates for Contrasts of Interest under π_{UU} (or π_{JU})

Contrasts	$\beta_3 - \beta_4$	$\beta_1 - \beta_3$	$\beta_1 - \beta_2$
B.E.	0.058	0.584	0.623
S.D.	0.282	0.207	0.320
$q_{.025}$	-0.501	0.165	0.064
$q_{.050}$	-0.409	0.234	0.234
$q_{.250}$	-0.126	0.451	0.401
$q_{.500}$	0.052	0.587	0.605
$q_{.750}$	0.237	0.720	0.826
$q_{.950}$	0.543	0.920	1.175
$q_{.975}$	0.646	0.988	1.299

3.7.2.3 Summary

- Goodness-of-fit of Bayesian procedure is demonstrated in both analyses. Based on these justifications, no evidence of lack-of-fit is revealed in either analysis.
- Priors π_{JU} and π_{UU} used in Analysis 2 are second order probability matching priors when $\beta_1, \beta_2, \beta_3, \beta_4$, or their differences are the parameters of interest. For γ_j 's, due to the discreteness of the underlying distribution, the matching concept cannot be justified rigorously. However, if one formally extends the matching concept from the continuous case to the discrete case, then π_{JU} is indeed a second order probability matching prior.
- We provide an example, where both nonidentifiability of the likelihood and censoring are present. The composite prior, nevertheless, provides a satisfactory matching performance for parameters of primary interest.

CHAPTER 4 GEOMETRIC COMPETING RISKS MODELS

4.1 Introduction

In Chapters 2 and 3, Bayesian inference was considered for competing risks models originating out of various BVE's. The nonidentifiability was overcome by the knowledge of the underlying model and prior belief. Suppose that the intrinsic bivariate lifetime mechanism is not clearly specified. One shall not interpret the result based on an *ad hoc* model assumption. In this chapter, we focus attention on a class of stochastic models for (T, Ψ, Δ) , which does not require exponential marginals or the independence of T and (Ψ, Δ) in advance. As we will see later, one also avoids nonidentifiability from such modeling scheme. In particular, we consider discretized lifetime T . This occurs for instance when lifetime T is screened periodically due to resource limitation or due to economic reason, sometimes even due to intentional discretization. The stochastic modeling scheme is stated in Section 4.2. Based on the specification of the stationary survival probabilities, generalized geometric competing risks models (GEM) are constructed, which feasibly feature a variety of lifetime phenomena, e.g., T has the piecewise exponential marginal. Interestingly enough, this approach reflects the mathematical evidence of the step function approximation to the true survival function. In Section 4.3, Bayesian inference under Laplace's and Jeffreys' priors are derived. The ECOG dataset, introduced in Section 4.4, is reanalyzed under the geometric model assumption.

4.2 Generalized Geometric Models and Likelihood Functions

Kaplan-Meier estimator or product-limit estimator (Kalbfleisch and Prentice, 1980 and Cox and Oakes, 1984) provides a stochastic modeling for analyzing univariate survival data. The survival curve is estimated by a step function under certain assumptions. Dirichlet process (Ferguson, 1973) for modeling the survival function, and Levy process (Kalbfleisch, 1978) or Beta process (Hjort, 1990) for modeling the underlying hazard function are some of the well-known Bayesian nonparametric alternatives. Clayton (1991), Aslanidou and Dey (1996), and Sinha and Dey (1997) extended these ideas to the analysis of multiple events, in which case the lifetime process is adjusted for covariates. This is also found convenient in the analysis of discretized survival data or interval censoring. We generalize this idea to establish a general class of competing risks models, which can in particular, feature the likelihood of the BVE without suffering from the nonidentifiability.

Let $0 = x_0 < x_1 < \cdots < x_{L+1} = \infty$, where $x_1, \dots, x_L$ denote the designed time points for survival screening or the intentional discretization of data in the study. Define $\tau_l = (x_{l-1}, x_l]$ for $1 \leq l \leq L$. Let $\mathbf{H}_l = (T^{(l)}, \Delta^{(l)}, \Psi^{(l)}; \mathbf{z})$ denote the lifetime history of (T, Δ, Ψ) in τ_l given $\mathbf{z}$ ($k \times 1$), where $T^{(l)}$ is the indicator of failure in τ_l , and $\Delta^{(l)}$ and $\Psi^{(l)}$ record the status of Δ and Ψ at τ_l , respectively. Define $\mathbf{H}_0 = (T^{(0)}, \Delta^{(0)}, \Psi^{(0)}; \mathbf{z}) \equiv (0, -1, -1; \mathbf{z})$. The three correlated stochastic processes are defined as follows.

$$T^{(l)} = \mathbf{1}_{[T \in \tau_l]},$$

$$\Psi^{(l)} = \begin{cases} 0 & \text{if } \Psi = 0, T^{(l)} = 1 \\ 1 & \text{if } \Psi = 1, T^{(l)} = 1 \\ -1 & \text{if } T^{(l)} = 0 \end{cases},$$

and

$$\Delta^{(l)} = \begin{cases} 0 & \text{if } \Delta = 0, \Psi^{(l)} = 1, T^{(l)} = 1 \\ 1 & \text{if } \Delta = 1, \Psi^{(l)} = 1, T^{(l)} = 1 \\ -1 & \text{if } \Psi^{(l)} = T^{(l)} = 0 \end{cases}.$$

$T^{(l)}$, $\Delta^{(l)}$, and $\Psi^{(l)}$ essentially comprise three monotone increasing stochastic processes. Define $0^0 \equiv 1$, $\mathbf{H}_l^+ = (T^{(l)} = 0, \Psi^{(l)}, \Delta^{(l)})$, and $\mathbf{H}_l^- = (T^{(l)} = 1, \Psi^{(l)}, \Delta^{(l)})$. Assume that $P(\mathbf{H}_0) = 1$, $P(T^{(l)} = 1 | \mathbf{H}_{l-1}^-) = 1$, $P(\Delta^{(l)} = -1, \Psi^{(l)} = -1 | T^{(l)} = 0, \mathbf{H}_{l-1}^+) = 1$, $P(\Psi^{(l)} = j | \Psi^{(l-1)} = j, \mathbf{H}_{l-1}^-) = 1$, and $P(\Delta^{(l)} = j | \Delta^{(l-1)} = j, \Psi^{(l-1)} = 1, \mathbf{H}_{l-1}^-) = 1$, for $j = 0, 1$, $1 \leq l \leq L$. The next table exhibits the possible transition routes of the joint process.

Table 4.1. Stochastic Transition Status

$\mathbf{H}_0$	$\mathbf{H}_{l-1}$	$\mathbf{H}_l$	Remark
$(0, -1, -1; \mathbf{z})$	$(0, -1, -1; \mathbf{z})$	$(0, -1, -1; \mathbf{z})$	Censored
	$(0, -1, -1; \mathbf{z})$	$(1, 0, -1; \mathbf{z})$	Simultaneous
	$(1, 0, -1; \mathbf{z})$	$(1, 0, -1; \mathbf{z})$	Failure
	$(0, -1, -1; \mathbf{z})$	$(1, 1, 0; \mathbf{z})$	Cause 2
	$(1, 1, 0; \mathbf{z})$	$(1, 1, 0; \mathbf{z})$	Failure
	$(0, -1, -1; \mathbf{z})$	$(1, 1, 1; \mathbf{z})$	Cause 1
	$(1, 1, 1; \mathbf{z})$	$(1, 1, 1; \mathbf{z})$	Failure

The likelihood of (t, δ, ψ) in terms of $(t^{(0)}, \psi^{(0)}, \delta^{(0)}), \dots, (t^{(L)}, \psi^{(L)}, \delta^{(L)})$ is given by

$$\begin{aligned} & P(\mathbf{H}_0) \left(\prod_{l=1}^L P(\mathbf{H}_l | \mathbf{H}_{l-1}) \right) \\ &= P(\mathbf{H}_0) \left(\prod_{l=1}^L P(t^{(l)} | \mathbf{H}_{l-1}) P(\psi^{(l)} | t^{(l)}, \mathbf{H}_{l-1}) P(\delta^{(l)} | \psi^{(l)}, t^{(l)}, \mathbf{H}_{l-1}) \right). \quad (4.1) \end{aligned}$$

The likelihood function is then determined by the transition probabilities $P(t^{(l)} | \mathbf{H}_{l-1}^+)$, $P(\psi^{(l)} | t^{(l)}, \mathbf{H}_{l-1}^+)$, and $P(\delta^{(l)} | \psi^{(l)}, t^{(l)}, \mathbf{H}_{l-1}^+)$. Let $F_t(\cdot)$, $F_\psi(\cdot)$, and $F_\delta(\cdot)$ be differentiable distribution functions, and these are given by

$$\begin{aligned}
F_t^{(l)} &= P(T^{(l)} = 1 | \mathbf{H}_{l-1}) = F_t(\mathbf{z}'\boldsymbol{\beta}_l; \tau_l) && \text{if } \mathbf{H}_{l-1} = \mathbf{H}_{l-1}^+, \\
&= P(T^{(l)} = 1 | \mathbf{H}_{l-1}) = 1 && \text{if } \mathbf{H}_{l-1} = \mathbf{H}_{l-1}^-; \quad (4.2)
\end{aligned}$$

$$\begin{aligned}
F_\psi^{(l)} &= P(\Psi^{(l)} = 1 | T^{(l)}, \mathbf{H}_{l-1}) = F_\psi(\mathbf{z}'\boldsymbol{\alpha}_l; \tau_l) && \text{if } T^{(l)} = 1, \mathbf{H}_{l-1} = \mathbf{H}_{l-1}^+, \\
&= P(\Psi^{(l)} = j | T^{(l)}, \mathbf{H}_{l-1}) = 1 && \text{if } \Psi^{(l-1)} = j, \mathbf{H}_{l-1} = \mathbf{H}_{l-1}^-; \quad (4.3)
\end{aligned}$$

$$\begin{aligned}
F_\delta^{(l)} &= P(\Delta^{(l)} = 1 | \Psi^{(l)}, T^{(l)}, \mathbf{H}_{l-1}) = F_\delta(\mathbf{z}'\boldsymbol{\gamma}_l; \tau_l) && \text{if } \Psi^{(l)} = T^{(l)} = 1, \mathbf{H}_{l-1} = \mathbf{H}_{l-1}^+, \\
&= P(\Delta^{(l)} = j | \Psi^{(l)}, T^{(l)}, \mathbf{H}_{l-1}) = 1 && \text{if } \Delta^{(l)} = j, \mathbf{H}_{l-1} = \mathbf{H}_{l-1}^-; \quad (4.4)
\end{aligned}$$

for $j = 0, 1$, $1 \leq l \leq L$. Let $\bar{F} = 1 - F$. The probabilities of nonsingularity and cause 1 failure by the end of the study are respectively given by

$$P(\Psi = 1 | T^{(L)} = 1) = \sum_{l=1}^L \{ (\prod_{0 \leq l' < l} \bar{F}_{l'}) F_t^{(l)} F_\psi^{(l)} \}, \quad (4.5)$$

$$P(\Delta = 1 | T^{(L)} = 1) = \sum_{l=1}^L \{ (\prod_{0 \leq l' < l} \bar{F}_{l'}) F_t^{(l)} F_\psi^{(l)} F_\delta^{(l)} \}. \quad (4.6)$$

Let $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_L)$, $\boldsymbol{\alpha}' = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_L)$, $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}'_1, \dots, \boldsymbol{\gamma}'_L)$, and $\mathbf{1}_s = \mathbf{1}_{[0 < P(\Psi=1) < 1]}$.

The joint likelihood function becomes

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{H}_0, \dots, \mathbf{H}_L) &= \prod_{l=1}^L \left\{ \left[F_t^{(l)} \left((F_\delta^{(l)})^{\delta^{(l)}} (\bar{F}_\delta^{(l)})^{1-\delta^{(l)}} \right)^{\psi^{(l)}} \right]^{\ell^{(l)}} (\bar{F}_t^{(l)})^{1-\ell^{(l)}} \right\} \\
&\times \left\{ \prod_{l=1}^L \left[(F_\psi^{(l)})^{\psi^{(l)}} (\bar{F}_\psi^{(l)})^{1-\psi^{(l)}} \right]^{\ell^{(l)}} \right\}^{\mathbf{1}_s}. \quad (4.7)
\end{aligned}$$

Under the noninformative censoring assumption, the marginal likelihood for censoring within τ_l is given by

$$\prod_{0 \leq l' < l} \bar{F}_t^{(l')}. \quad (4.8)$$

Denote a generalized geometric variable Y by $Y \sim \text{Geo}(L; p_1, \dots, p_L)$ if it has pdf of the form

$$\begin{aligned} P(Y = i) &= \left(\prod_{l'=0}^l (1 - p_{l'}) \right) p_{l+1} \quad \text{for } 0 \leq l < L, \\ &= \prod_{l'=1}^L (1 - p_{l'}) \quad \text{for } l = L, \end{aligned} \quad (4.9)$$

where $p_0 = 0$ and $0 < p_l < 1$ for $1 \leq l \leq L$. Clearly, $\sum_{l=1}^m P(Y = l) = 1$. Moreover, (4.9) presents a variant of a regular geometric distribution function, which is subject to truncation and different “success” probabilities. By this definition, one obtains

$$\sum_{l=1}^L \mathbf{1}_{[T^{(l)}=0]} = \sum_{l=1}^L \mathbf{1}_{[\Psi^{(l)}=-1]} = \sum_{l=1}^L \mathbf{1}_{[\Delta^{(l)}=-1]} \sim \text{Geo}(L; F_t^{(1)}, \dots, F_t^{(L)}). \quad (4.10)$$

(4.2)-(4.4) are thus referred to as the generalized geometric competing risks model (GEM). The truncated geometric model proposed by Mori, Woollen, and Woolworth (1995) is actually a special case of this model, where the transition probability function is the logistic distribution.

This generalized geometric model with appropriate specification of F ’s allows for characterizing a variety of bivariate lifetime structures. For instance, if T follows the piecewise exponential distribution (Breslow, 1974), then $F_t^{(l)}(\mu_l) = 1 - e^{-\mu_l(x_l - x_{l-1})}$. This family of distributions broadly encompasses many lifetime distributions via its smooth mathematical approximation. Suppose that $\beta_1 = \dots = \beta_L$, $\alpha_1 = \dots = \alpha_L$, and $\gamma_1 = \dots = \gamma_L$. Particularly, let $x_1, \dots, x_L$ be chosen evenly. It leads to a simpler version of geometric models with “equal transition probabilities”, i.e.,

$$F_t \equiv F_t^{(l)}, \quad F_\psi \equiv F_\psi^{(l)}, \quad F_\delta \equiv F_\delta^{(l)}, \quad (4.11)$$

for $1 \leq l \leq L$. The likelihood function under (4.11) is then given by

$$L(\beta, \alpha, \gamma | \mathbf{H}_0, \dots, \mathbf{H}_L)$$

$$= \bar{F}_t^w(F_t)^{\mathbf{1}_{[w < L]}} \prod_{l=1}^L \left\{ \left[(F_\delta)^{\delta^{(l)}} (\bar{F}_\delta)^{1-\delta^{(l)}} \right]^{\psi^{(l)}} \left[(F_\psi)^{\psi^{(l)}} (\bar{F}_\psi)^{1-\psi^{(l)}} \right]^{\mathbf{1}_s} \right\}^{t^{(l)}}, \quad (4.12)$$

where w is the realization of $W = \sum_{l=1}^L \mathbf{1}_{[T^{(l)}=0]}$ with $W \sim \text{Geo}(L, F_t, \dots, F_t)$. And

$$P(\Psi = 1 | T^{(L)} = 1) = \sum_{l=1}^L \{(\bar{F}_t)^{l-1} F_t F_\psi\}, \quad (4.13)$$

$$P(\Delta = 1 | T^{(L)} = 1) = \sum_{l=1}^L \{(\bar{F}_t)^{l-1} F_t F_\psi F_\delta\}. \quad (4.14)$$

Further, suppose that $F_t(y) = 1 - e^{-y}$ and $F_\psi(y) = F_\delta(y) = e^y/(1 + e^y)$. Then a family of singular BVE's is obtained, where the continuous waiting time T follows Cox's proportional hazard model, i.e., $T \sim G(1, e^{\mathbf{x}'\boldsymbol{\beta}})$ and

$$\begin{aligned} \mathbf{1}_{[\Psi^{(l)}=1|T^{(l)}=1]} &\sim \text{Bin}\left(1, \frac{e^{\mathbf{x}'\boldsymbol{\alpha}}}{1 + e^{\mathbf{x}'\boldsymbol{\alpha}}}\right), \\ \mathbf{1}_{[\Delta^{(l)}=1|T^{(l)}=\Psi^{(l)}=1]} &\sim \text{Bin}\left(1, \frac{e^{\mathbf{x}'\boldsymbol{\gamma}}}{1 + e^{\mathbf{x}'\boldsymbol{\gamma}}}\right). \end{aligned}$$

Consequently, under $P(\mathbf{1}_s = 1) = 1$, (4.7) matches the generalized ACBVE likelihood function derived in Chapter 3 in the absence of the indicator multiplier.

Next, we derive noninformative priors and the corresponding posteriors based on (4.7) and (4.12).

4.3 Bayesian Analysis

Laplace's prior and Jeffreys' prior are employed in the Bayesian analysis.

4.3.1 Likelihood Function

Let $T_i^{(l)}$, $\Psi_i^{(l)}$, and $\Delta_i^{(l)}$ denote the indicators of failure, nonsingularity, and cause 1 failure in τ_l for individual i , $1 \leq i \leq n_+$. Suppose that $(t_i^{(l)}, \psi_i^{(l)}, \delta_i^{(l)})$ is the realization

of $(T_i^{(l)}, \Psi_i^{(l)}, \Delta_i^{(l)})$, respectively. Let failures are indexed by $1, \dots, n$, and individuals who have earlier dropout before x_L are indexed by $n+1, \dots, n_+$. Define $R_0 = \{1, \dots, n_+\}$, and for $1 \leq l \leq L$, $R_l = \{i : t_i^{(l)} = 0, 1 \leq i \leq n_+\}$, $\tilde{R}_l = \{i : i \in R_{l-1} \text{ and } t_i^{(l)} = 1, 1 \leq i \leq n\}$, and $\tilde{R}_l^{(1)} = \{i : i \in \tilde{R}_l \text{ and } \Psi_i^{(l)} = 1\}$. Denote $F_{t,i}^{(l)} = F_t(\mathbf{z}'_i \boldsymbol{\beta}_l)$, $F_{\psi,i}^{(l)} = F_\psi(\mathbf{z}'_i \boldsymbol{\alpha}_l)$, and $F_{\delta,i}^{(l)} = F_\delta(\mathbf{z}'_i \boldsymbol{\gamma}_l)$. Under (4.2)-(4.4), and (4.8), the joint likelihood function for $\{(t_i, \delta_i, \psi_i) : 1 \leq i \leq n_+\}$ is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{d}) = L_t(\boldsymbol{\beta} | \mathbf{d}) L_\psi(\boldsymbol{\alpha} | \mathbf{d}) L_\delta(\boldsymbol{\gamma} | \mathbf{d}), \quad (4.15)$$

where

$$L_t(\boldsymbol{\beta} | \mathbf{d}) \propto \prod_{l=1}^L \left\{ \left(\prod_{i \in \tilde{R}_l} F_{t,i}^{(l)} \right) \prod_{i \in R_l} (\bar{F}_{t,i}^{(l)}) \right\}, \quad (4.16)$$

$$L_\psi(\boldsymbol{\alpha} | \mathbf{d}) \propto \prod_{l=1}^L \left\{ \prod_{i \in \tilde{R}_l^{(1)}} (F_{\psi,i}^{(l)})^{\psi_i^{(l)}} (\bar{F}_{\psi,i}^{(l)})^{1-\psi_i^{(l)}} \right\}^{\mathbf{1}_n}, \quad (4.17)$$

$$L_\delta(\boldsymbol{\gamma} | \mathbf{d}) \propto \prod_{l=1}^L \left\{ \prod_{i \in \tilde{R}_l^{(1)}} (F_{\delta,i}^{(l)})^{\delta_i^{(l)}} (\bar{F}_{\delta,i}^{(l)})^{1-\delta_i^{(l)}} \right\}. \quad (4.18)$$

The posteriors associated with Jeffreys' prior and Laplace's prior are calculated as follows.

4.3.2 Laplace's Prior and the Posterior

Suppose that Laplace uniform prior is assumed. The posterior density is proportional to (4.15), i.e.,

$$\pi_U = \pi_U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \propto 1, \quad (4.19)$$

$$\pi_U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{d}) \propto L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{d}). \quad (4.20)$$

Apparently, the posterior marginals are: $\pi_U(\boldsymbol{\beta}|\mathbf{d}) \propto L_t(\boldsymbol{\beta}|\mathbf{d})$, $\pi_U(\boldsymbol{\alpha}|\mathbf{d}) \propto L_\psi(\boldsymbol{\alpha}|\mathbf{d})$, and $\pi_U(\boldsymbol{\gamma}|\mathbf{d}) \propto L_\delta(\boldsymbol{\gamma}|\mathbf{d})$. Let $A \subset \{1, \dots, n_+\}$, and $\mathbf{Z}'_A$ denote the sub-matrix of $\mathbf{Z}'$ consisting of column vectors of $\mathbf{Z}'$ with indices belonging to A . Using Shao and Chen's (1998) results, the propriety of $\pi_U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{d})$ holds if all the following requirements are met.

$$(C.4.1) \quad \mathbf{Z}_{\tilde{R}_l^{(1)}} \text{ is of rank } k, \text{ for } 1 \leq l \leq L \text{ since } R_{l-1} \supseteq \tilde{R}_l \supseteq \tilde{R}_l^{(1)}.$$

$$(C.4.2) \quad \text{Let } \mathbf{U}_{R_l}, \mathbf{U}_{\tilde{R}_l}, \text{ and } \mathbf{U}_{\tilde{R}_l^{(1)}} \text{ be the diagonal matrices with the } i^{\text{th}} \text{ diagonal elements given by } (-1)^{\epsilon_i^{(l)}} \text{ for } i \in R_l, (-1)^{\phi_i^{(l)}} \text{ for } i \in \tilde{R}_l, \text{ and } (-1)^{\delta_i^{(l)}} \text{ for } i \in \tilde{R}_l^{(1)}, \text{ respectively. Then it requires that } [\mathbf{Z}_{R_l}]' \mathbf{U}_{R_l} \mathbf{v}_t^{(l)} = \mathbf{0}, [\mathbf{Z}_{\tilde{R}_l}]' \mathbf{U}_{\tilde{R}_l} \mathbf{v}_\phi^{(l)} = \mathbf{0}, \text{ and } [\mathbf{Z}_{\tilde{R}_l^{(1)}}]' \mathbf{U}_{\tilde{R}_l^{(1)}} \mathbf{v}_\delta^{(l)} = \mathbf{0}, \text{ for some positive vectors } \mathbf{v}_t^{(l)}, \mathbf{v}_\phi^{(l)} \text{ and } \mathbf{v}_\delta^{(l)} \text{ for } 1 \leq l \leq L.$$

$$(C.4.3) \quad \int_{-\infty}^{\infty} |u|^k dF_t(u) < \infty, \int_{-\infty}^{\infty} |u|^k dF_\psi(u) < \infty, \text{ and } \int_{-\infty}^{\infty} |u|^k dF_\delta(u) < \infty.$$

Under (4.11) and Laplace's prior, the joint posterior is given by

$$\pi_U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{d}) \propto \pi_U(\boldsymbol{\beta}|\mathbf{d}) \pi_U(\boldsymbol{\alpha}|\mathbf{d}) \pi_U(\boldsymbol{\gamma}|\mathbf{d}), \quad (4.21)$$

where

$$\pi_U(\boldsymbol{\beta}|\mathbf{d}) \propto \prod_{l=1}^L \left\{ \prod_{i \in R_l} (F_{t,i}) \prod_{i \in \tilde{R}_l} (\bar{F}_{t,i}) \right\}, \quad (4.22)$$

$$\pi_U(\boldsymbol{\alpha}|\mathbf{d}) \propto \prod_{l=1}^L \prod_{i \in \tilde{R}_l} \left\{ (F_{\psi,i})^{\psi_i^{(l)}} (\bar{F}_{\psi,i})^{(1-\psi_i^{(l)})} \right\}^{\mathbf{1}_s}, \quad (4.23)$$

$$\pi_U(\boldsymbol{\gamma}|\mathbf{d}) \propto \prod_{l=1}^L \prod_{i \in \tilde{R}_l^{(1)}} \left\{ (F_{\delta,i})^{\delta_i^{(l)}} (\bar{F}_{\delta,i})^{(1-\delta_i^{(l)})} \right\}. \quad (4.24)$$

Let $\tilde{R}^{(1)} = \{i : i \in \bigcup_{l=1}^L \tilde{R}_l^{(1)}\}$ and $\tilde{R} = \{i : i \in \bigcup_{l=1}^L \tilde{R}_l\}$. Then besides (C.4.3), the propriety of (4.21) is ensured under the following assumptions.

$$(C.4.4) \quad \mathbf{Z}_{\tilde{R}^{(1)}} \text{ satisfies the full column rank assumption since } \tilde{R} \supseteq \tilde{R}^{(1)}.$$

(C.4.5) Let $\mathbf{U}_{n_+}$, $\mathbf{U}_{\tilde{R}}$, $\mathbf{U}_{\tilde{R}^{(1)}}$ be the diagonal matrices with the i^{th} diagonal elements given by $(-1)^{\mathbf{1}_{\{i \geq n_+\}}}$ for $1 \leq i \leq n_+$, $(-1)^{\phi_i}$ for $i \in \tilde{R}$ and $(-1)^{\delta_i}$ for $i \in \tilde{R}^{(1)}$, respectively. The propriety requires that $[\mathbf{Z}]'\mathbf{U}_{n_+}\mathbf{v}_t = \mathbf{0}$, $[\mathbf{Z}_{\tilde{R}}]'\mathbf{U}_{\tilde{R}}\mathbf{v}_\phi = \mathbf{0}$, and $[\mathbf{Z}_{\tilde{R}^{(1)}}]'\mathbf{U}_{\tilde{R}^{(1)}}\mathbf{v}_\delta = \mathbf{0}$, for some positive vectors $\mathbf{v}_t$, $\mathbf{v}_\phi$ and $\mathbf{v}_\delta$.

4.3.3 Jeffreys' Prior and the Posterior

Let $f^{(l)}(y) = \partial F^{(l)}(y)/\partial y$. Define three diagonal matrices $\mathbf{M}_t^{(l)}$ ($n_+ \times n_+$), $\mathbf{M}_\psi^{(l)}$ ($n \times n$), and $\mathbf{M}_\delta^{(l)}$ ($n \times n$), and their i^{th} diagonal elements are accordingly given by

$$M_{t,i}^{(l)} \equiv M_{t,i}^{(l)}(\beta) = \left(\frac{f_{t,i}^2}{F_{t,i}^{(l)} \bar{F}_{t,i}^{(l)}} \right) \left\{ \prod_{l' < l} \bar{F}_{t,i}^{(l')} \right\}, \quad (4.25)$$

$$M_{\psi,i}^{(l)} \equiv M_{\psi,i}^{(l)}(\beta, \alpha) = \left(\frac{f_{\psi,i}^2}{F_{\psi,i}^{(l)} \bar{F}_{\psi,i}^{(l)}} \right) \left\{ \left(\prod_{l' < l} \bar{F}_{t,i}^{(l')} \right) F_{t,i}^{(l)} \right\}, \quad (4.26)$$

$$M_{\delta,i}^{(l)} \equiv M_{\delta,i}^{(l)}(\beta, \alpha, \gamma) = \left(\frac{f_{\delta,i}^2}{F_{\delta,i}^{(l)} \bar{F}_{\delta,i}^{(l)}} \right) \left\{ \left(\prod_{l' < l} \bar{F}_{t,i}^{(l')} \right) F_{t,i}^{(l)} F_{\psi,i}^{(l)} \right\}. \quad (4.27)$$

Let $\mathbf{M}_t = \oplus [\mathbf{Z}'\mathbf{M}_t^{(1)}\mathbf{Z}, \dots, \mathbf{Z}'\mathbf{M}_t^{(L)}\mathbf{Z}]$, $\mathbf{M}_\psi = \oplus [\mathbf{Z}_n'\mathbf{M}_\psi^{(1)}\mathbf{Z}_n, \dots, \mathbf{Z}_n'\mathbf{M}_\psi^{(L)}\mathbf{Z}_n]$, and $\mathbf{M}_\delta = \oplus [\mathbf{Z}_n'\mathbf{M}_\delta^{(1)}\mathbf{Z}_n, \dots, \mathbf{Z}_n'\mathbf{M}_\delta^{(L)}\mathbf{Z}_n]$. The Fisher information matrix is given by

$$\mathbf{I}(\beta, \alpha, \gamma) = \oplus [\mathbf{M}_t, \mathbf{M}_\psi, \mathbf{M}_\delta] \text{ (or } \oplus [\mathbf{M}_t, \mathbf{M}_\delta]), \quad (4.28)$$

which depends on the assumption of singularity. Applying the same technique used in Chapter 3, Jeffreys' prior is calculated as

$$\begin{aligned} \pi_J(\beta, \alpha, \gamma) &\propto \prod_{l=1}^L \left\{ \left(\sum_{P^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \prod_{u=1}^k M_{t,i_u}^{(l)} \right) \right. \\ &\quad \times \left. \left(\sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \prod_{u=1}^k M_{\delta,i_u}^{(l)} \right) \right\} \end{aligned}$$

$$\times \left(\sum_{P_n^{(k)}} c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) \prod_{u=1}^k M_{\psi, i_u}^{(l)} \right)^{\mathbf{1}_s \frac{1}{2}}, \quad (4.29)$$

where $P^{(k)} = \{(i_1, \dots, i_k) : c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) > 0, 1 \leq i_1 < i_2 < \dots < i_k \leq n_+\}$ and $P_n^{(k)} = \{(i_1, \dots, i_k) : c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k}) > 0, 1 \leq i_1 < i_2 < \dots < i_k \leq n\}$. We shall prove in the next theorem that the Jeffreys' prior given in (4.29) is indeed a proper pdf provided that $\mathbf{Z}_n$ is of rank k .

Theorem 4.1 Assume that $\mathbf{Z}_n$ is a full rank design matrix. Let $\pi_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ be given in (4.29), then

$$\int \dots \int \pi_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) d\boldsymbol{\beta} d\boldsymbol{\alpha} d\boldsymbol{\gamma} < \infty.$$

Proof Since $0 \leq \bar{F} \leq 1$, one obtains the inequalities

$$M_{\cdot, i}^{(l)} \leq \frac{(f_{\cdot, i}^{(l)})^2}{F_{\cdot, i}^{(l)} \bar{F}_{\cdot, i}^{(l)}}. \quad (4.30)$$

Applying Jensen's inequality, one has

$$\begin{aligned} & \pi_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \\ & \leq \prod_{l=1}^L \left\{ \sum_{P^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})]^{\frac{1}{2}} \left(\prod_{u=1}^k \frac{f_{t, i_u}^{(l)}}{(F_{t, i_u}^{(l)} \bar{F}_{t, i_u}^{(l)})^{\frac{1}{2}}} \right) \right\} \\ & \times \prod_{l=1}^L \left\{ \sum_{P_n^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})]^{\frac{1}{2}} \left(\prod_{u=1}^k \frac{f_{\delta, i_u}^{(l)}}{(F_{\delta, i_u}^{(l)} \bar{F}_{\delta, i_u}^{(l)})^{\frac{1}{2}}} \right) \right\} \\ & \times \prod_{l=1}^L \left\{ \sum_{P_n^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})]^{\frac{1}{2}} \left(\prod_{u=1}^k \frac{f_{\psi, i_u}^{(l)}}{(F_{\psi, i_u}^{(l)} \bar{F}_{\psi, i_u}^{(l)})^{\frac{1}{2}}} \right) \right\}^{\mathbf{1}_s}. \end{aligned}$$

Using the change-of-variables $y_{i_u}^{(l)} = \mathbf{z}'_{i_u} \boldsymbol{\beta}_l$, for $(i_1, \dots, i_k) \in P^{(k)}$, the propriety of $\pi_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ follows from the finiteness of the beta integral

$$\int \frac{f(y)}{(F(y)\bar{F}(y))^{\frac{1}{2}}} dy = (\Gamma(.5))^2 = \pi. \quad (4.31)$$

Next, we consider the reduced model given in (4.11). Jeffreys' prior is given by

$$\begin{aligned} & \pi_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \\ & \propto \left\{ \sum_{P^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})] \prod_{u=1}^k \left(\frac{f_{t,i_u}^2}{(F_{t,i_u} \bar{F}_{t,i_u})} \right) \left(\sum_{l=1}^L (\bar{F}_{t,i_u})^{l-1} \right) \right\}^{\frac{1}{2}} \\ & \times \left\{ \sum_{P_n^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})] \prod_{u=1}^k \left[\frac{f_{\delta,i_u}^2}{(F_{\delta,i_u} \bar{F}_{\delta,i_u})} \left(\sum_{l=1}^L (\bar{F}_{t,i_u})^{l-1} F_{t,i_u} \right) F_{\psi,i_u}^{\mathbf{1}_{i_u}} \right] \right\}^{\frac{1}{2}} \\ & \times \left\{ \sum_{P_n^{(k)}} [c(\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_k})] \prod_{u=1}^k \left[\frac{f_{\psi,i_u}^2}{(F_{\psi,i_u} \bar{F}_{\psi,i_u})} \left(\sum_{l=1}^L (\bar{F}_{t,i_u})^{l-1} F_{t,i_u} \right) \right]^{\mathbf{1}_s} \right\}^{\frac{1}{2}}. \quad (4.32) \end{aligned}$$

Since $\sum_{l=1}^L (\bar{F}_{t,i})^{l-1} < L$ and $\sum_{l=1}^L (\bar{F}_{t,i})^{l-1} F_{t,i} < 1$, $\int \pi_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) d\boldsymbol{\beta} d\boldsymbol{\alpha} d\boldsymbol{\gamma}$ is proper according to Theorem 4.1. The corresponding posterior is thus proper if $\mathbf{Z}_n$ is of rank k .

4.4 Application to a Sample with Categorical Covariates

As cited in Section 3.5, when a sample is finitely classified into k categories, the parameters associated with category j are independent of those associated with any other category. Thus, x_i 's may be independently assigned for each category. Let L_j denote the number of time grids for category j , i.e., the cut-off time points are $0 = x_{j_0} < x_{j_1} < \dots < x_{j_{L_j}} < x_{j_{L_j+1}} = \infty$. Let $\boldsymbol{\beta}'_j = (\beta_j^{(1)}, \dots, \beta_j^{(L_j)})$, $\boldsymbol{\alpha}'_j = (\alpha_j^{(1)}, \dots, \alpha_j^{(L_j)})$, and $\boldsymbol{\gamma}'_j = (\gamma_j^{(1)}, \dots, \gamma_j^{(L_j)})$ be the parameters associated with category

j , $1 \leq j \leq k$. Denote $\beta' = (\beta'_1, \dots, \beta'_k)$, $\alpha' = (\alpha'_1, \dots, \alpha'_k)$, and $\gamma' = (\gamma'_1, \dots, \gamma'_k)$. Define $n_{1_j}^{(l)} = \sum_{i \in \bar{R}_i^{(l)}} \psi_i^{(l)} \delta_i^{(l)} \mathbf{1}_{[z_i, j=1]}$, $n_{2_j}^{(l)} = \sum_{i \in \bar{R}_i^{(l)}} \psi_i^{(l)} (1 - \delta_i^{(l)}) \mathbf{1}_{[z_i, j=1]}$, $n_{12_j}^{(l)} = \sum_{i \in \bar{R}_i^{(l)}} (1 - \psi_i^{(l)}) \mathbf{1}_{[z_i, j=1]}$, $n_{\cdot j}^{(l)} = n_{1_j}^{(l)} + n_{2_j}^{(l)} + n_{12_j}^{(l)}$, and $\bar{n}_{\cdot j}^{(l)} = \sum_{i \in \bar{R}_i^{(l)}} (1 - t_i^{(l)}) \mathbf{1}_{[z_i, j=1]}$. Denote $F_{\beta_j}^{(l)} = F_t(\beta_j^{(l)})$, $F_{\alpha_j}^{(l)} = F_{\psi}(\alpha_j^{(l)})$, and $F_{\gamma_j}^{(l)} = F_{\delta}(\gamma_j^{(l)})$. The likelihood function becomes

$$\begin{aligned} L(\beta, \alpha, \gamma | \mathbf{d}) &= \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(F_{\beta_j}^{(l)})^{n_{\cdot j}^{(l)}} (\bar{F}_{\beta_j}^{(l)})^{\bar{n}_{\cdot j}^{(l)}} (F_{\gamma_j}^{(l)})^{n_{1_j}^{(l)}} (\bar{F}_{\gamma_j}^{(l)})^{n_{2_j}^{(l)}} \right] \right\} \\ &\times \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(F_{\alpha_j}^{(l)})^{n_{1_j}^{(l)} + n_{2_j}^{(l)}} (\bar{F}_{\alpha_j}^{(l)})^{n_{12_j}^{(l)}} \right]^{1_s} \right\}. \end{aligned} \quad (4.33)$$

Accordingly, Jeffreys' prior is given by

$$\begin{aligned} \pi_J(\beta, \alpha, \gamma) &\propto \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(\bar{F}_{\beta_j}^{(l)})^{L_j - l - \frac{1}{2}} (F_{\gamma_j}^{(l)} \bar{F}_{\gamma_j}^{(l)})^{-\frac{1}{2}} f_{\beta_j}^{(l)} f_{\gamma_j}^{(l)} \right] \right\} \\ &\times \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(\bar{F}_{\beta_j}^{(l)})^{\frac{L_j - l}{2}} (F_{\beta_j}^{(l)})^{\frac{1}{2}} (\bar{F}_{\alpha_j}^{(l)})^{-\frac{1}{2}} f_{\alpha_j}^{(l)} \right]^{1_s} \right\}. \end{aligned} \quad (4.34)$$

The posteriors under Laplace's prior and Jeffreys' prior are respectively given by $\pi_U(\beta, \alpha, \gamma | \mathbf{d}) \propto L(\beta, \alpha, \gamma | \mathbf{d})$ and

$$\begin{aligned} &\pi_J(\beta, \alpha, \gamma | \mathbf{d}) \\ &\propto \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(\bar{F}_{\beta_j}^{(l)})^{\bar{n}_{\cdot j}^{(l)} + L_j - l - \frac{1}{2}} (F_{\beta_j}^{(l)})^{n_{\cdot j}^{(l)}} f_{\beta_j}^{(l)} \right] \right\} \\ &\times \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(F_{\gamma_j}^{(l)})^{n_{1_j}^{(l)} - \frac{1}{2}} (\bar{F}_{\gamma_j}^{(l)})^{n_{2_j}^{(l)} - \frac{1}{2}} f_{\gamma_j}^{(l)} \right] \right\} \\ &\times \prod_{j=1}^k \left\{ \prod_{l=1}^{L_j} \left[(\bar{F}_{\beta_j}^{(l)})^{\frac{L_j - l}{2}} (F_{\beta_j}^{(l)})^{\frac{1}{2}} (F_{\alpha_j}^{(l)})^{n_{1_j}^{(l)} + n_{2_j}^{(l)}} (\bar{F}_{\alpha_j}^{(l)})^{n_{12_j}^{(l)} - \frac{1}{2}} f_{\alpha_j}^{(l)} \right]^{1_s} \right\}. \end{aligned} \quad (4.35)$$

The criteria for the propriety of $\pi_U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{d})$ are $n_{1_j}^{(l)} > 0$ and $n_{2_j}^{(l)} > 0$ under the nonsingular model assumption. $n_{12_j}^{(l)} > 0$ is required additionally for the singular model.

4.5 Data Analysis

The ECOG lung cancer example is revisited in this Section. We consider the sample categorization defined in Section 3.7. The time grids for each category are chosen separately. A natural way to choose the time grids is based on the Kaplan-Meier scheme, distinct failure times being chosen as time interval ends. In this example, the joint posterior is improper when Laplace's prior is assumed since some of $n_{1_j}^{(l)}$'s or $n_{2_j}^{(l)}$'s are zeros. We thus consider only the posteriors associated with Jeffreys' prior. No particular forms of $F_{\cdot}^{(l)}$'s are specified. We focus on the inference of $F_{\cdot}^{(l)}$'s. Indeed, the posterior distributions of $\beta_j^{(l)}$'s and $\gamma_j^{(l)}$'s can be obtained via one-to-one transformation from $F_t^{(l)}$'s and $F_{\delta}^{(l)}$'s, respectively. The likelihood of ACBVE's can be viewed as that of a GEM with one grid (cf. Section 4.2). We compare the results derived under the two model assumptions. Table 4.2 presents posterior means and standard deviations of cause 1 failure probabilities under the ACBVE and the GEM.

Table 4.2. Comparison of Posterior Means and Standard Deviations for $P(\Delta = 1|T^{(L)} = 1)$ under the ACBVE and the GEM

(X_1, X_2)	Model	B.E.	S.D.
(1, 1)	ACBVE	.6220	.6264
	GEM	.5514	.1510
(1, 0)	ACBVE	.7308	.5481
	GEM	.5770	.1938
(0, 1)	ACBVE	.6639	.6667
	GEM	.5853	.1061
(0, 0)	ACBVE	.5938	.6056
	GEM	.5298	.2132

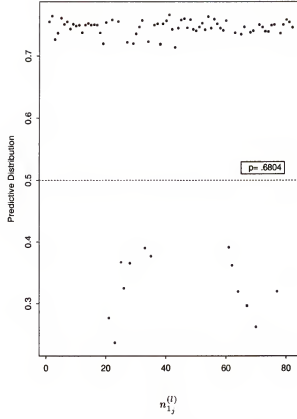


Figure 4.1. Bayesian Goodness-of-Fit Procedure for $n_{1j}^{(l)}$'s: plotting posterior predictive matching probabilities of $n_{1j}^{(l)}$'s.

The Bayesian goodness-of-fit procedure for $n_{1j}^{(l)}$'s is presented by the scatter plot of $p_{n_{1j}^{(l)}}$ given in Figure 4.1, where $p_{n_{1j}^{(l)}}$ is computed from (3.51).

It suggests that the geometric model is preferable in terms of prediction (cf. Figure 3.4). From Table 4.3, the posterior standard deviations of $P(\Delta = 1|T^{(L)} = 1)$'s are substantially smaller under the GEM in comparison with the ACBVE. It thus, evidences that the stochastic modeling enhances the precision of estimate.

The estimated survival curves under the two models are drawn along with the data empirical quantiles of $\prod_{t'=1}^l \bar{F}_t^{(t')}$ in Figure 4.2.

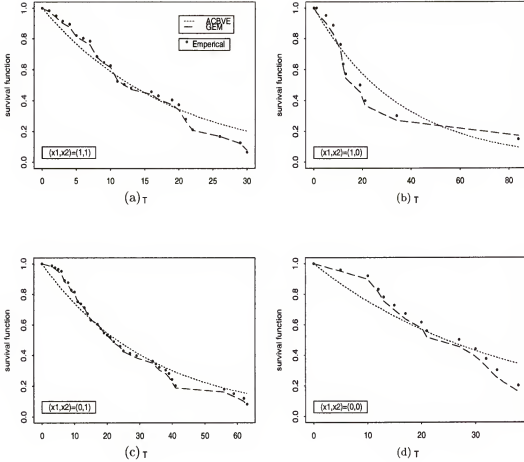


Figure 4.2. Survival Curves under the ACBVE and the GEM in Comparison with the Data Empirical Quantiles: (a) estimated survival curves for $(X_1, X_2) = (1, 1)$; (b) estimated survival curves for $(X_1, X_2) = (1, 0)$; (c) estimated survival curves for $(X_1, X_2) = (0, 1)$; (d) estimated survival curves for $(X_1, X_2) = (0, 0)$.

It reveals consistence between the estimated survival probabilities under the GEM assumption and the observed data, which again supports the advantage of incorporating Bayesian method with this modeling strategy.

CHAPTER 5

SUMMARY AND FUTURE RESEARCH

Competing risks models often lead to nonidentifiable likelihoods. The resulting inferential problem can be resolved by employing Bayesian methodology with informative priors (cf. Section 2.7) or stagewise noninformative priors (cf. Chapters 2 and 3), where the underlying models are assumed to be BVE's, or utilizing a more general class of stochastic models for the competing risks outcomes (cf. Chapter 4). In the simulation study demonstrated in Section 2.7, we find that with the use of informative priors, Bayesians successfully guide the inference of even nonidentifiable parameters. The present research, however, focuses primarily on Bayesian method with noninformative priors. The proposed stagewise priors are appealing due to various probability matching properties. In Chapter 4, a class of flexible competing risks models is proposed, which unlike BVE's, does not require a specific model assumption. Also, even without any model specification, when covariates are categorical, Laplace's and Jeffreys' priors lead to beta posteriors for the transition probabilities. Moreover, for analysis of a particular dataset, precision of estimates is enhanced by this flexible modeling.

The mirror side of competing risks is referred to as “complementary risks”, where the data consist of $T^c = \max(T_1, T_2)$, and $I^c = j\mathbf{1}_{[T_3-j \leq T_j]}$, for instance, a parallel system of two components. The likelihood of (T^c, I^c) once again is not fully identifiable. As in our case, this nonidentifiability can be viewed as a consequence of data incompleteness. Incorporating Bayesian methodology into such cases can result in fruitfully inferential procedures.

APPENDIX
PROOFS OF MATCHING PROPERTIES OF π_{JU} AND π_J

First order probability matching

$$\sum_{j=1}^k \frac{\partial}{\partial \theta_j} [(I^{11})^{-1/2} I^{i1} \pi(\boldsymbol{\theta})] = 0,$$

since I^{uv} 's and $\pi(\boldsymbol{\theta})$ are free from $\boldsymbol{\theta}$.

Second order probability matching

We first evaluate $\tau^{jr} = I^{j1} I^{r1} / I^{11}$ and $L_{jru} = E_{\boldsymbol{\theta}} \{ \partial^3 l(\boldsymbol{\theta}) / \partial \theta_j \partial \theta_r \partial \theta_u \}$, $1 \leq r, u \leq$

k . Clearly, τ^{jr} and $L_{jru} = \sum_{i=1}^k n_{\cdot i} a_{ij} a_{ir} a_{iu}$ are all independent of $\boldsymbol{\theta}$. Thus

$$\begin{aligned} \frac{\partial^2}{\partial \theta_j \partial \theta_r} \{ \tau^{jr} \pi(\boldsymbol{\theta}) \} &= 0 \\ \frac{\partial}{\partial \theta_v} \{ L_{jru} \tau^{jr} (3I^{uv} - 2\tau^{uv}) \pi(\boldsymbol{\theta}) \} &= 0. \end{aligned}$$

HPD matching

Since

$$\begin{aligned} L_{j,ru} &= E_{\boldsymbol{\theta}} \{ \partial l(\boldsymbol{\theta}) / \partial \theta_j \partial^2 l(\boldsymbol{\theta}) \partial \theta_r \partial \theta_u \} \\ &= E_{\boldsymbol{\theta}} \left\{ \left(\sum_{i=1}^k n_{\cdot i} a_{ij} - a_{ij} e^{\mathbf{a}'_i \boldsymbol{\theta} t_{s_i}} \right) \left(\sum_{i=1}^k n_{\cdot i} a_{ir} a_{iu} \right) \right\} \\ &= 0 \end{aligned}$$

and $\partial \pi_{JU}(\boldsymbol{\theta}) / \partial \theta_r = \partial \pi_J(\boldsymbol{\theta}) / \partial \theta_r = 0$, the HPD property is satisfied.

REFERENCES

- Armitage, P. and Zippin, C. (1966). Use of Concomitant Variables and Incomplete Survival Information in the Estimation of an Exponential Survival Parameter. *Biometrics*, **22**, 665-72.
- Arnold, B. C. (1975). Multivariate Exponential Distributions Based on Hierarchical Successive Damage. *Journal of Applied Probability*, **12**, 142-7.
- Aslanidou, H. and Dey, D. K. (1996). Model Determination for Multivariate Survival Data under Gamma Frailty Models. *ASA Proceedings of the Biometrics Section*, 292-7.
- Basu, A. P. and Ghosh, J. K. (1978). Identifiability of the Multinomial and Other Distributions under Competing Risks Model. *Journal of Multivariate Analysis*, **8**, 413-29.
- Basu, A. P. and Ghosh, J. K. (1980). Identifiability of Distributions under Competing Risks and Complementary Risks Model. *Communications Statistics, Part A - Theory and Method*, **14**, 1515-25.
- Berger, J. O. and Bernardo, J. M. (1989). Estimating a Product of Means: Bayesian Analysis with Reference Priors. *Journal of the American Statistical Association*, **84**, 200-7.
- Berger, J. O. and Sun D. (1993). Bayesian Analysis for the Poly-Weibull Distribution. *Journal of the American Statistical Association*, **87**, 1412-18.
- Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference (with discussion). *Journal of the Royal Statistical Society, Series B*, **41**, 113-28.
- Block, H. W. and Basu, A. P. (1974). A Continuous Bivariate Exponential Extension. *Journal of the American Statistical Association*, **69**, 1031-7.
- Breslow, N. (1974). Covariance Analysis of Censored Survival Data. *Biometrics*, **30**, 89-99.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167-74.
- Chiang, C. L. (1970). Competing Risks and Conditional Probabilities. *Biometrics*, **26**, 767-76.
- Chib, S. and Greenberg E. (1995). Regression Models and Lifetables. *The American Statistician*, **49**, 327-36.
- Clayton, D. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models. *Biometrics*, **47**, 467-85.

- Clayton, D. and Cuzick, J. (1985). Multivariate Generalizations of Proportional Hazards Model. *Journal of the Royal Statistical Society, Series A*, **48**, 82-117.
- Cox, D. R. (1972). Regression Models and Lifetables. *Journal of the Royal Statistical Society, Series B*, **34**, 189-220.
- Datta, G. S. (1995). On Priors Proving Frequentist Validity for Bayesian Inference for Multiple Parametric Functions. *Biometrika*, **82**, 37-45.
- Datta, G. S. and Ghosh, M. (1995). Some Remarks on Noninformative Priors. *Journal of the American Statistical Association*, **90**, 1357-63.
- David H. A. and Moeschberger, M. L. (1978). The Theory of Competing Risks. New York: Macmillan.
- Dellaportas, P. and Smith, F. M. (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. *Applied Statistics*, **42**(3), 443-59.
- Dey, D. K. and Mukerjee, R. (1993). Frequentist Validity of Posterior Quantiles in the Presence of a Nuisance Parameter: Higher Order Asymptotics. *Biometrika*, **80**, 499-505.
- Downton, F. (1970). Bivariate Exponential Distributions in Reliability Theory. *Journal of the Royal Statistical Society, Series B*, **33**, 408-17.
- Feigl, P. and Zelen, M. (1965). Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics*, **21**, 826-38.
- Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, **1**, 209-30.
- Freund, J. E. (1961). A Bivariate Extension of the Exponential Distribution. *Journal of the American Statistical Association*, **56**, 971-7.
- Friday, D. S. and Patil, G. P. (1977). A Bivariate Exponential Model with Applications to Reliability and Computer Generation of Random Variables. Theory and Applications of Reliability, 1, eds.. New York: Academic Press.
- Gelman, A., Carlin, J., Stern, H. S., and Rubin, D. (1996). Applied Statistical Science. Bayesian Data Analysis. London: Chapman and Hall.
- Gelman, A. and Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences *Statistical Science*, **7**(4), 457-511.
- Ghosh, J. K. and Mukerjee, R. (1993). Frequentist Validity of Highest Posterior Density Regions in the Presence of Nuisance Parameters. *Statistics and Decisions*, **13**, 131-39.
- Ghosh, M. and Mukerjee, R. (1996). Recent Developments on Probability Matching Priors. Applied Statistical Science, **3**, Commack, NY: Nova Science Publishers.
- Gumbel, E. J. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association*, **56**, 698-707.

- Hjort, N. L. (1990). Nonparametric Bayesian Estimators Based on Beta Process in Models for Lifetime History Data. *The Annals of Statistics*, **18**, 1259-94.
- Hougaard, P. (1986). A Class of Multivariate Failure Time Distributions. *Biometrika*, **73**, 671-8.
- Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian Analysis of Generalized Linear Models Using Jeffreys's Prior. *Journal of the American Statistical Association*, **86**, 981-6.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press.
- Kalbfleisch, J. D. (1978). Nonparametric Bayesian Analysis of Survival Time Data. *Journal of the Royal Statistical Association, Series B*, **40**, 214-21.
- Kalbfleisch, J. D. and Mackay, R. J. (1979). On Constant-Sum Models for Censored Survival Data. *Biometrika*, **66**, 87-90.
- Kalbfleisch, J. D. and Prentice, P. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Klein, J. P. and Basu, A. P. (1982). Accelerated Life Tests under Competing Weibull Causes of Failure. *Communications in Statistics, Theory and Methods*, **11**(20), 2271-86.
- Lagakos, S. W. (1978). A Covariate Models for Partially Censored Data Subject to Competing Causes of Failures. *Applied Statistics*, **27**, 235-41.
- Laplace, P. (1812). *Theorie Analytique des Probabilities*. Paris: Courcier.
- Lee, L. (1979). Multivariate Distributions Having Weibull Properties. *Journal of Multivariate Analysis*, **9**, 267-77.
- Lehmann, E. L. (1966). Some Concepts of Dependence. *Annals of Mathematical Statistics*, **37**, 1137-53.
- Liu, Z. Z. and Nakao, Z. (1990). On Bayesian Parameters and Reliability Estimation in a Bivariate Exponential Model. *Japonica*, **35**, 949-57.
- Liu, Z. Z. and Nakao, Z. (1991). Empirical Bayesian Interval Estimation for Reliability Function of a Bivariate Exponential Model. *Japonica*, **37**, 1085-91.
- Lu, J. C. (1989). Weibull Extensions of the Freund and Marshall-Olkin Bivariate Exponential Models. *IEEE Transactions on Reliability*, **38**, 615-9.
- Lu, J. C. (1992). Bayes Parameter Estimation for the Bivariate Weibull Model of Marshall-Olkin for Censored Data. *IEEE Transactions on Reliability*, **41**(4), 608-20.
- Lu, J. C. and Bhattacharyya, G. K. (1990). Some New Constructions of Bivariate Weibull Models. *Annals of the Institute of Statistical Mathematics*, **42**, 543-59.
- Lu, J. C. and Bhattacharyya, G. K. (1991). Inference Procedures for a Bivariate Exponential Model of Gumbel Based on Life Test of Component and System. *Journal of Statistical Planning and Inference*, **27**, 383-96.

- Marshall, A. W. and Olkin, I. (1967). A Multivariate Exponential Distribution. *Journal of the American Statistical Association*, **62**, 30-44.
- McCullagh, P. and Nelder, J. A. (1996). Generalized Linear Models. London: Chapman and Hall.
- Moeschberger, M. L. (1971). Lifetests under Competing Causes of Failure and the Theory of Competing Risks. *Biometrics*, **27**, 909-33.
- Moeschberger, M. L. (1974). Life Tests under Competing Causes of Failure. *Technometrics*, **16**(1), 39-46.
- Mori, M., Woolson, R. F., and Woodworth, G. G. (1995). Slope Estimation in the Presence of Informative Right Censoring: Modeling the Number of Observations as a Geometric Random Variable, *Biometrics*, **50**, 39-50.
- Noble B. and James W. D. (1977). Applied Linear Algebra. Englewood Cliffs. NJ: Prentice Hall.
- Ryu, K. (1993). An Extension of Marshall and Olkin's Bivariate Exponential Distribution. *Journal of the American Statistical Association*, **87**, 1458-65.
- Sarkar, S. K. (1987). A Continuous Bivariate Exponential Distribution. *Journal of the American Statistical Association*, **82**, 667-75.
- Shao, Q. M. and Chen, M. H. (1998). Propriety of Posterior Distribution for Dichotomous Quantal Response Models. *Unpublished manuscript*.
- Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association*, **92**, 1195-1212.
- Tibshirani, R. (1989). Noninformative Priors for One Parameter of Many. *Biometrika*, **76**, 604-8.
- Tsiatis, A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks. *National Academy of Sciences*, **72**, 20-2.
- Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, **16**, 439-54.
- Wada, C. Y., Sen, P. K., and Shimakura, S. E. (1996). A Bivariate Exponential Model with Covariates in Competing Risks Data. *Calcutta Statistical Association Bulletin* **46**, 197-210.
- Welch, B. and Peers., H. W. (1963). On Formulae for Confidence Points Based on Integrals of Weighted Likelihoods. *Journal of the Royal Statistical Society, Series B*, **24**, 318-29.
- Williams, J. S. and Lagakos, S. W. (1977). Models for Censored Survival Analysis: Constant-Sum and Variable-Sum Models. *Biometrika*, **64**, 215-22.

BIOGRAPHICAL SKETCH

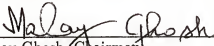
Chen-Pin Wang was born on December 6, 1970, in Chungli, Taiwan, to J. C. Wang and H. Y. Chen. Along with her elder brother, Jeng-Peng, and younger brother, Jeng-Han, this family was complete.

Music was Chen-Pin's favorite subject in childhood. She sought to study piano before her interest in science showed itself during her high school years. Following her graduation from high school, Chen-Pin pursued mathematics as her major at the Central University in Chungli, Taiwan. Her journey as a student upon would not have been fulfilled without the absolute support and encouragement of her most beloved father, J. C. Wang, who died in 1991. Chen-Pin was fortunate to be able to endure the loss and to continue searching and the pursuit of her lifetime goals. Inspired by Dr. Y. S. Chow, professor of statistics at the University of Columbia, she decided to explore the world of statistics.

Following her graduation, Chen-Pin moved to Florida in 1993 and enrolled in a Ph.D. degree program at the University of Florida, Department of Statistics.

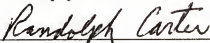
Following her graduation in August, 1999, Chen-Pin will relocate to Tampa, Florida, to further her academic career at the University of South Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



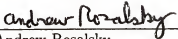
Malay Ghosh, Chairman
Distinguished Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



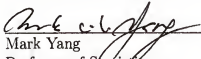
Randolph Carter
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Andrew Rosalsky
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Mark Yang
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Yunmei Chen
Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1999

Dean, Graduate School